

# Climate change and sustainability: an actuarial risk management perspective

Xu, Yanbin

2024

Xu, Y. (2024). Climate change and sustainability: an actuarial risk management perspective. Doctoral thesis, Nanyang Technological University, Singapore.  
<https://hdl.handle.net/10356/180071>

<https://hdl.handle.net/10356/180071>

<https://doi.org/10.32657/10356/180071>

---

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

*Downloaded on 22 Oct 2024 23:08:09 SGT*

# Climate Change and Sustainability: An Actuarial Risk Management Perspective

Yanbin Xu

Nanyang Business School

A thesis submitted to the Nanyang Technological University  
in partial fulfillment of the requirement for the degree of  
Doctor of Philosophy

2024

## Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

01/05/2024

---

Date

ITU NTU NTU NTU NTU NTU NTU NTU  
NTU NTU NTU NTU NTU NTU NTU NTU  
ITU NTU NTU NTU NTU NTU NTU NTU  
ITU NTU NTU NTU NTU NTU NTU NTU

*Yanbin Xu*

---

Yanbin Xu



## Authorship Attribution Statement

This thesis contains material from three papers accepted at conferences, where I am listed as an author.

Chapter 2 was presented at the 24th International Congress on Insurance: Mathematics and Economics (2021) and the 19th China International Conference of Insurance and Mathematics (2021).

The contributions of the co-authors are as follows:

- I proposed the idea, processed the analysis, and wrote the manuscript.
- The manuscript was revised under the supervision of Prof Guangming Pan, Prof Ken Seng Tan, and Prof Wenjun Zhu.

Chapter 3 was presented at Presented at the 58th Actuarial Research Conference (2023) and the 26th Insurance: Mathematics and Economics International Conference (2023).

The contributions of the co-authors are as follows:

- I proposed the idea, processed the analysis, and wrote the manuscript.
- The manuscript was revised under the supervision of Prof Ken Seng Tan and Prof Wenjun Zhu.

Chapter 4 was presented at Presented at the 27th Insurance: Mathematics and Economics International Conference (2024).

The contributions of the co-authors are as follows:

- I proposed the idea, proved the propositions and theorems, and wrote the manuscript.
- The manuscript was revised under the supervision of Prof Shinichi Kamiya, Prof Lysa Porth, and Prof Prof Wenjun Zhu.

01/05/2024

Date

ITU NTU NTU NTU NTU NTU NTU NTU  
NTU NTU NTU NTU NTU NTU NTU NTU  
ITU NTU NTU NTU NTU NTU NTU NTU  
ITU NTU NTU NTU NTU NTU NTU NTU

Yanbin Xu

Yanbin Xu

## Acknowledgements

I am profoundly grateful for the support and guidance of my supervisors, Prof. Wenjun Zhu, Prof. Lysa Porth, and Prof. Ken Seng Tan. Over the past four years, they have guided me to the foothills of the vast academic mountain. This thesis and all the work I have completed during my candidature would not exist without their invaluable help. Their insights and encouragement were crucial in shaping this work, and I am deeply thankful for their patience and dedication.

I would also like to extend my sincere gratitude to my Thesis Advisory Committee, Prof. Tao Chen and Prof. Benjamin P. Horton, as well as my Examination Committee, Prof. Hwang Byoung-Hyoun, Prof. Luo Jiang, and Prof. Wu Yuan. Their rigorous feedback and constructive criticism have been invaluable in refining this thesis.

Special thanks are due to Prof. Jun-Koo Kang, Prof. Chuan-Yang Hwang, Prof. Chang Xin (Simba), Prof. Guangming Pan, Prof. Shinichi Kamiya, Prof. Jinggong Zhang, Prof. Pengyu Wei, Dr. Ciyu Nie (Jade), Prof. Pingyi Lou, Dr. Adam Shao, Prof. Joey Zhou, Prof. Bowen Yang for their support, valuable discussions, and the high quality of academic training they provided. This has enriched my understanding and expanded my perspective on actuarial science, finance, statistics, and economics.

I am thankful to the Global-Asia Insurance Partnership (GAIP), who generously funded my research and study. Their support has been instrumental in achieving the goals of this project.

I extend my heartfelt gratitude to my family, who have provided unwavering support and encouragement throughout my academic journey. To my parents, Mr. Dong Xu and Ms. Yan Tian, who have always believed in me—your faith in my abilities has been a constant source of strength. To my grandfather, Mr. Liren Xu, your wisdom and value system will continue to guide me throughout my life. May you rest in peace.

I am also immensely grateful to my fiancée, Ms. Yifan Wu, whose love, understanding, and patience have been a pillar of stability in my life. Thank you for being there at every step of this journey.

I would like to acknowledge some of the renowned leaders and experts in the finance and insurance industry, with whom I had the privilege to work and learn: Ms. Angela Jiang and Ms. Qing Yu (Martha) from Franklin Templeton, Ms. Pei Nee Goi from Miller Insurance Services, Dr. Qingyao Yu from Allianz Reinsurance, Mr. Qin Sun, Mr. Sagar Sahay, and Mr. Ivan Zheng from Munich Re, Ms. Xincheng Cai (Kandice) from PetSure, Ms. Pay Peng Neo from ActuariesOnTap, Ms. Keira Liang from AIA, and Ms. Guoyu Huang (Adela) from UBS. Their camaraderie and willingness to share their knowledge have greatly enhanced my

understanding of the financial and insurance industry. The practitioner perspectives I gained from them have tremendously shaped my research interests and focus.

I also owe my thanks to Ms. Valerie De Souza, Ms. Quek Bee Hua, and Ms. Karen Barlaan for their excellent work in providing PhD students at NBS with the best working environment during our candidature.

Finally, I extend my appreciation to everyone who was involved directly or indirectly in the completion of this thesis. Your contributions have not gone unnoticed, and I am thankful for your part in my journey.

## **Dedication**

Dedicated to the memory of Professor Ken Seng Tan, my supervisor and mentor, whose wisdom and encouragement profoundly shaped my academic journey. His legacy will continue to inspire future generations. May his kindness and wisdom be forever remembered.

# Table of Contents

Statement of Originality	ii
Supervisor Declaration Statement	iii
Authorship Attribution Statement	iv
Acknowledgements	v
Dedication	vii
List of Figures	xiv
List of Tables	xvi
Abstract	xviii
<b>1 Introduction</b>	<b>1</b>
<b>2 A Sustainable Area-yield Insurance Program with Optimal Risk Pooling: A Behavior-based Spectral Clustering Approach</b>	<b>5</b>
2.1 Introduction . . . . .	5
2.2 Area-yield Insurance Program . . . . .	12
2.3 Sustainable Risk Pooling Framework . . . . .	14

2.3.1	Determining the Optimal Number of Clusters: A Behavior-based Approach . . . . .	15
2.3.2	Optimal Risk Pooling Algorithm . . . . .	19
2.4	Empirical Analysis . . . . .	21
2.4.1	Data . . . . .	21
2.4.2	Baseline Results of the Sustainable Risk Pooling Framework . . . . .	24
2.4.3	Subsample Analysis: Geographical Implications . . . . .	28
2.4.4	Performances of the Gap Statistic-based Framework . . . . .	31
2.5	Robustness of Sustainable Risk Pooling . . . . .	32
2.5.1	Impact of Sustainable Compliance Ratio . . . . .	33
2.5.2	Impact of Retained Variance Ratio . . . . .	33
2.5.3	Impact of Cost Ratio . . . . .	35
2.5.4	Alternative Measures of Basis Risk . . . . .	35
2.5.5	Optimal Area-yield Contracts . . . . .	36
2.5.6	Constructing Risk Pools with the Same Size . . . . .	38
2.6	A Farm-level Analysis Based on Simulated Data . . . . .	39
2.6.1	Farm-level Data Simulation . . . . .	39
2.6.2	Performance of Sustainable Risk Pooling . . . . .	40
2.7	Conclusion Remarks . . . . .	41
2A	Appendix . . . . .	43
2A.1	Proofs . . . . .	43
2A.2	Additional analyses . . . . .	45

<b>3</b>	<b>From Meteorology to Market: A Geo-Hierarchical Deep Learning Approach for Flood Risk Pricing</b>	<b>48</b>
3.1	Introduction . . . . .	48
3.2	A Geo-Hierarchical Deep Learning Model . . . . .	55
3.2.1	Model Architecture . . . . .	56
3.2.2	Convolutional Neuron Network Model . . . . .	58
3.2.3	Wavelet Neuron Network Model . . . . .	60
3.3	Performance in Flood Risk Prediction . . . . .	64
3.3.1	Data . . . . .	65
3.3.2	Performance Metrics . . . . .	69
3.3.3	Model Selection for Flood Risk Prediction . . . . .	71
3.3.4	GHDL Performance in Flood Risk Forecasting . . . . .	73
3.4	Performance in Flood Insurance Pricing . . . . .	76
3.4.1	National Flood Insurance Program . . . . .	76
3.4.2	Flood Insurance Pricing . . . . .	77
3.4.3	GLM Model Summary . . . . .	79
3.4.4	Net Premium and Solvency Capital Requirement . . . . .	80
3.5	Interpretability of the GHDL Model . . . . .	84
3.5.1	Relative Importance Analysis . . . . .	84
3.5.2	Partial Dependence Analysis . . . . .	85
3.6	Conclusion Remarks . . . . .	85
<b>4</b>	<b>Bridging the Protection Gap: A Tax Redistribution Solution Under a Private-Public Partnership Framework</b>	<b>88</b>
4.1	Introduction . . . . .	88

4.2	The Tax Redistribution Model . . . . .	94
4.2.1	Public Disaster Relief Liability . . . . .	95
4.2.2	Insurers' Incentive . . . . .	95
4.2.3	Residents' Incentive . . . . .	97
4.2.4	Redistribution . . . . .	98
4.2.5	Planer's Problem . . . . .	101
4.3	Insurance with Extreme Loss . . . . .	102
4.3.1	Risk . . . . .	103
4.3.2	Insurance Demand Function . . . . .	104
4.3.3	Insurance Supply Function . . . . .	108
4.3.4	Optimal Redistribution . . . . .	111
4.4	Conclusion Remarks . . . . .	113
4A	Appendix . . . . .	114
4A.1	Proofs . . . . .	114
<b>5</b>	<b>Resilience Through Responsibility: The Impact of ESG Status on Firm Resilience to Climate-Induced Disruptions</b>	<b>120</b>
5.1	Introduction . . . . .	120
5.2	Potential Channels of Resilience Improvements . . . . .	123
5.2.1	Market Sentiment . . . . .	124
5.2.2	Financial Stability Proxy . . . . .	125
5.2.3	Climate Risk Hedging . . . . .	125
5.3	Methodology . . . . .	126
5.3.1	Data . . . . .	126
5.3.2	Key Variables Definition . . . . .	126

5.3.3	Empirical Methodology . . . . .	132
5.4	Results . . . . .	134
5.4.1	Impact of Hurricane Events on Stock Returns . . . . .	134
5.4.2	ESG and Climate Resilience . . . . .	137
5.4.3	Natural and Unnatural ESG Firms . . . . .	139
5.4.4	The Impact of Being ESG Enthusiasts . . . . .	144
5.4.5	Long-term ESG Status and Fundamentals after Hurricane Shocks . . . . .	146
5.5	Robustness tests . . . . .	150
5.5.1	ESG Endogeneity . . . . .	150
5.5.2	Climate Resilience in the Most Impacted Industries . . . . .	152
5.6	Conclusion Remarks . . . . .	153
5A	Appendix . . . . .	156
5A.1	Variable Definitions . . . . .	156
5A.2	Calculation of Cumulative Abnormal Returns . . . . .	156
5A.3	Natural Catastrophes Used in RDD . . . . .	157
<b>6</b>	<b>Conclusion and Future Works</b>	<b>161</b>
6.1	Conclusion Remarks . . . . .	161
6.2	Future Work Directions . . . . .	162
6.2.1	Alternative Dependence Structure for Area-yield Pooling . . . . .	162
6.2.2	GHDL Model Application to Emerging Market . . . . .	163
6.2.3	General Externality in Tax Redistribution Modeling . . . . .	163
6.2.4	Optimal Public Sector Participation Rate . . . . .	164
6.2.5	Releasing Assumptions in Insurers' Financing and Investing . . . . .	164
6.2.6	General Class of Green Assets' Climate Resilience . . . . .	164
6.2.7	Green Assets Climate Resilience Channel Testing . . . . .	164

References	166
Glossary	184

# List of Figures

2.1	Heartland region and U.S. corn production (Illustrative period of 2015-19). . .	22
2.2	The ratios of fixed and variable costs to value of corn produced for heartland region. The grey bars highlight the census years. <i>Source:</i> USDA Economic Research Service. . . . .	25
2.3	Agricultural District of the heartland region and the proposed sustainable risk pooling results. . . . .	27
2.4	Illinois geography and risk pooling visualization. . . . .	30
2.5	Basis risk of area-yield contracts using risk pools derived from various numbers of principal components. . . . .	34
2.6	Basis risk improvement in the sustainable risk pooling framework compared to AGD, measured against changes in fixed and variable costs. . . . .	36
2.7	Distribution of the performance metrics from different risk pooling methods on simulated farm-level data. . . . .	41
3.1	Spatio-temporal flood risk correlation across a cluster of six cities in the Mississippi River basin . . . . .	52
3.2	Information integration process of GHDL structure . . . . .	57
3.3	Geographic connectivity around Kansas city . . . . .	58
3.4	The GHDL structure for Kansas city . . . . .	59
3.5	Illustration of padding, stride, convolution, and pooling procedure . . . . .	60

3.6	Coverage of the time-frequency plane of the wavelet and Fourier based transforms . . . . .	63
3.7	A 3-level discrete wavelet transform system . . . . .	64
3.8	The cluster of six cities the Mississippi river basin . . . . .	66
3.9	The GHDL structure . . . . .	67
3.13	Net premium vs. Loss . . . . .	82
4.1	Taxation on per policy written in low risk regions . . . . .	99
4.2	Description of tax redistribution in high and low risk regions . . . . .	100
4.3	Coverage demand curve of both region under constant absolute risk aversion utility. . . . .	109
5.1	Number of firms covered by MSCI ESG database by year. . . . .	127
5.2	Lagged correlation of ESG scores between year $t$ and year $t + k(k = 1$ to $4)$ . . . . .	128
5.3	Impact regions of hurricane irma. . . . .	131
5.4	Geographical distribution of Compustat firms' firm-year headquarters at the county level from 2003 to 2020. . . . .	131
5.5	Natural and unnatural ESG industries, classified by their excess ESG scores. . . . .	142
5.6	ESG enthusiasts in natural and unnatural ESG industries. . . . .	145

# List of Tables

2.1	Descriptive Statistics . . . . .	23
2.2	Baseline Results: Sustainable Risk Pooling Quality Analysis . . . . .	29
2.3	Performances of the Gap Statistic-Based Risk Pooling . . . . .	32
2.4	Robustness Check: Impact of Sustainable Compliance Ratio Selection . . . . .	34
2.5	Robustness Check: Alternative Basis Risk Measures . . . . .	37
2.6	Robustness Check: Impact of Protection <i>scale</i> . . . . .	38
2.7	Robustness Check: Impact of a Same-Size Constraint . . . . .	39
2A.1	Impact of High Dimensionality . . . . .	46
2A.2	Sustainable Risk Pooling Results with Parametric Utilities . . . . .	47
3.1	Selected Cities and their Corresponding Counties and Stations . . . . .	66
3.2	Geo-Hierarchical Deep Learning Model Selection for Flood Risk Forecasting . . . . .	72
3.3	Performance in Flood Risk Forecasting . . . . .	75
3.4	Descriptive Statistics of the NFIP . . . . .	77
3.5	General Linear Model Result in Flood Risk Pricing . . . . .	81
3.6	Performance in Flood Insurance Pricing . . . . .	83
4.1	Joint Probability Structure of HR Risk . . . . .	104
4.2	Discrete States Framework . . . . .	105

5.1	Correlation Matrix of ESG, E, S, and G Scores . . . . .	128
5.2	Hurricanes Utilized for Analyses: Characteristics and Impact . . . . .	130
5.3	Summary Statistics . . . . .	133
5.4	Univariate Analysis of Cumulative Abnormal Returns: Assessing Hurricane Impact on Impacted and Non-Impacted Firms . . . . .	136
5.5	Univariate Analysis of Cumulative Abnormal Returns: Assessing Hurricane Impact on High and Low ESG Score Firms . . . . .	138
5.6	Baseline Analysis: The Influence of ESG Status and Hurricane Impact on Firms	140
5.7	Natural and Unnatural ESG Firms' Resilience to Hurricane Impact . . . . .	143
5.8	ESG Enthusiasts' Resilience to Hurricane Impact . . . . .	147
5.9	Impact of Firms' ESG Status and Hurricane Shocks on Long-term ESG Status and Fundamentals . . . . .	149
5.10	Regression Discontinuity Analysis: The Impact of ESG Policy Adoption on Firm Resilience to Natural Catastrophe Shocks . . . . .	151
5.11	Top Five Industries Most Impacted by Hurricane Strikes . . . . .	152
5.12	Robustness Analysis: The Impact of ESG Status on Climate Resilience in the Most Impacted Industries . . . . .	154
5A.1	Variable Definitions . . . . .	156
5A.2	Natural Catastrophes Utilized for RDD: Characteristics and Impact . . . . .	157

## Abstract

Climate risk poses significant challenges to socio-economic sustainability, yet the financial and insurance industries generally remain reactive, hindered by a limited understanding of climate impacts and inadequate modeling techniques. This thesis aims to study climate risk from an actuarial science perspective, enhancing our capabilities in modeling, mitigating, and adapting to these changes.

The first project in this thesis examines the agricultural industry, which is significantly impacted by climate change. I introduce a behavior-based machine learning approach to optimize risk pooling in area-yield insurance, addressing challenges such as moral hazard, high administrative costs, and data sparsity. By analyzing farming behavior under area-yield insurance contracts through a utility maximization framework and employing unsupervised spectral clustering, this study effectively reduces basis risk and enhances the sustainability of insurance programs.

The second project proposes a solution for the intensifying risk of climate-related floods. It develops a geo-hierarchical deep learning model for flood risk assessment that does not rely on high-resolution or hard-to-access data, making it particularly suitable for emerging markets with data limitation. This model aims to refine actuarial practices for flood insurance, improving predictive accuracy and economic efficiency.

Another challenge posed by climate risk is its disproportionate impact, which has widened the protection gap in highly exposed regions. The third project discusses a self-financing tax redistribution scheme under a private-public partnership framework to manage this climate risk-related protection gap across different risk regions. The model assesses the externalities involved in wealth transfers between moderate and high-risk areas, proposing solutions to mitigate negative impacts through effective policy interventions.

While the previous three projects focus on studying climate risk on the liability side of insurance companies, the final project extends the discussion to the asset side. I find that green assets exhibit stronger valuation resilience during the natural catastrophe events, underscoring the benefit of including them in the portfolios of insurance and other financial institutions. After addressing the endogeneity concerns, I identify market sentiment as the channel of such benefit.

# Chapter 1

## Introduction

In recent decades, human activities have significantly altered the Earth's climate, leading to what is now known as anthropogenic climate change ([Chen et al., 2021](#)). Industrialization, deforestation, and the burning of fossil fuels have resulted in unprecedented increases in greenhouse gas (GHG) emissions. These activities have intensified the natural greenhouse effect, causing global warming ([IPCC, 2021](#)). As a result, we are experiencing rising global temperatures, melting ice caps, sea level rise, and an increasing frequency and intensity of extreme weather events such as hurricanes, heatwaves, and droughts ([Fox-Kemper et al., 2021](#); [Lin et al., 2016](#); [Saintilan et al., 2020](#); [Mukherjee et al., 2018](#)). While natural processes like volcanic eruptions and variations in solar radiation can influence climate, the current trend is predominantly driven by human influence ([IPCC, 2021](#)).

Looking ahead, future predictions of climate change suggest a continuing trend of global warming with significant impacts on both natural and human systems. According to climate models, average global temperatures are projected to rise by 1.5 to 4.5 degrees Celsius by the end of the 21st century, depending on the trajectory of GHG emissions ([Lee et al., 2021](#)). This warming is expected to lead to more frequent and severe climate-related natural hazards, particularly affecting vulnerable populations ([Benevolenza and DeRigne, 2019](#); [Helldén et al., 2021](#); [Rocque et al., 2021](#)).

However, despite the robust evidence supporting climate change, there are uncertainties in predicting its precise future impacts. These uncertainties arise from the complexity of

climate systems, the variability of natural processes, and the limitations of current climate models (Hawkins and Sutton, 2009; Collins et al., 2013). Such uncertainty are particularly important on fragmatic regions and small islands like South Eastern Asia (Arias et al., 2021). Key uncertainties include the sensitivity of the climate system to GHG concentrations, the potential for climate feedback mechanisms, and the regional variability of climate impacts (Flato et al., 2014; Sherwood et al., 2020). Additionally, future socio-economic developments, policy decisions, and technological advancements will influence the trajectory of GHG emissions, adding another layer of unpredictability (Chen et al., 2021). Navigating these uncertainties is crucial for developing effective strategies that enhance resilience and mitigate the risks associated with climate change.

The escalating frequency and severity of climate-related natural hazards underscore a critical and growing concern for global economic stability and sustainable development. In 2022 alone, economic losses from such hazards soared to over USD 270 billion, highlighting not only the profound financial repercussions but also the urgent sustainability challenges posed by climate change (Swiss Re, 2024). This surge in losses not only exemplifies the direct impacts on infrastructure and economies but also underscores the broader environmental and social imperatives that necessitate sustainable solutions. If current trends continue, annual direct losses could exceed USD 400 billion by the 2030s, emphasizing the need for comprehensive and integrated risk management strategies that foster both economic resilience and environmental sustainability (Swiss Re, 2024).

Beyond economic impacts, climate change also poses serious challenges to human health and life expectancy. According to the World Health Organization, climate change is expected to cause an additional 250,000 deaths annually between 2030 and 2050 due to malnutrition, malaria, and heat stress (WHO, 2023). Additionally, vulnerable groups, including the elderly, are increasingly at risk of diseases intensified by climate changes (Romanello et al., 2022). These health impacts make it more difficult to achieve the Sustainable Development Goals set by the United Nations, especially those related to eliminating poverty and hunger (UN, 2023).

The importance of developing innovative and effective risk management strategies grows

increasingly evident. Traditional approaches to risk assessment are being continuously tested by the unprecedented scale and unpredictability of climate impacts. This necessitates a shift towards more dynamic, informed, and resilient risk management frameworks. Actuarial science, with its rigorous focus on risk quantification and financial implications, stands at the forefront of this shift, offering vital tools and methodologies for navigating the uncertainties that lie ahead. The urgency of climate change demands that actuarial science not only adapts but also innovates, thus ensuring that risk management strategies not only protect socio-economic interests but also foster a sustainable future.

This thesis investigates how actuarial techniques can enhance sustainability efforts and address climate-related risks. By integrating advanced computational models with traditional actuarial practices, and exploring potential policy tools and sustainable investment opportunities, this work aims to broaden the contribution of actuaries to environmental sustainability and economic resilience.

Chapters 2 and 3 of this thesis focus on enhancing the capacity for climate risk modelling. In Chapter 2, I examine the impact of climate risk on a crucial and highly vulnerable industry: agriculture. I propose that area-yield insurance, as opposed to individual yield insurance, offers a promising alternative for hedging production risks associated with climate change due to its potential to reduce moral hazard. This chapter introduces a behavior-based machine learning approach to create risk groupings, significantly improving the basis risk associated with area-yield insurance products.

Chapter 3 addresses the challenges climate change poses to the sustainability of flood insurance. I present a geo-hierarchical deep learning (GHDL) model that not only integrates climate and physical data into actuarial models for flood insurance pricing but does so more cost-effectively and robustly than typical physical risk models.

In addition to modeling challenges, climate change disproportionately affects various regions and sectors, thereby widening the protection gap significantly. Chapter 4 addresses this issue by proposing a self-financing tax redistribution framework aimed at mitigating the expansion of the protection gap. This chapter also provides an implicit solution for optimal redistribution.

The challenges posed by climate change extend beyond the liabilities of insurers to include their asset sides as well. Identifying assets with strong climate resilience is crucial for insurers because it reduces the risk of insolvency, particularly in the event of catastrophic events. In Chapter 5, using hurricane strikes as natural experiments, I have determined that green assets exhibit greater climate resilience, with confirmed causality.

Lastly, Chapter 6 provides concluding remarks and discusses future research directions.

Through a series of focused studies, this thesis demonstrates how actuarial practices can be evolved to meet the challenges posed by climate change, thereby supporting sustainability through more precise risk assessment and management. Each chapter delves into a specific area where actuarial science can intersect meaningfully with public policy, insurance, and corporate governance to mitigate the effects of environmental risks, close protection gaps, and enhance societal resilience.

# Chapter 2

## A Sustainable Area-yield Insurance Program with Optimal Risk Pooling: A Behavior-based Spectral Clustering Approach

### 2.1 Introduction

In recent years, the concept of sustainability has gained significant traction across various sectors and industries. The United Nations has adopted the Sustainable Development Goals (SDGs) as a critical framework to mitigate environmental impact and foster long-term resilience. Agriculture is the world's largest industry, and its sustainability is of vital environmental, social, and economic importance. However, climate disasters pose significant challenges to food security, which are transmitted through the global interconnected agricultural supply chain (Nagurney et al., 2024). Agricultural insurance is a prominent risk management tool to help producers reduce production uncertainty and mitigate financial risk (Iturrioz, 2009; Mahul and Stutley, 2010; Porth and Tan, 2015). In recent years, we have witnessed a rapid growth of the agricultural insurance sector. In 2019, the global crop insurance market size was estimated to be \$34.05 billion, and is projected to reach \$53.02

billion by 2027 (Goswami et al., 2020). The U.S. Federal Crop Insurance Program (FCIP), which has provided Multiple Peril Crop Insurance policies since its establishment in 1938, remains the largest agricultural insurance market in the world. For example, in 2019, the total liability of the U.S. FCIP exceeded \$116 billion which is equivalent to more than a quarter of the value of agricultural production (Rosch, 2021).

However, traditional indemnity insurance programs are faced with several challenges that hinder their sustainability. First, there can be significant information asymmetry issues, including moral hazard and adverse selection (Goodwin, 1993; Miranda and Glauber, 1997; Rowell and Connelly, 2012; Hao et al., 2018). Second, assessments of individual farm-level losses can be lengthy and costly. Large underwriting loss and high administrative costs increase the premiums and reduce the participation rate, and hence, these insurance programs rely heavily on government subsidies.<sup>1</sup> Third, the successful implementation of crop insurance programs is further hampered by the limited and unreliable nature of farm-level data (Woodard, 2016; Zhu et al., 2019).

Area-based insurance schemes, prominently featuring the Area Risk Protection Insurance (ARPI) in the U.S., present a viable alternative to traditional indemnity-based programs. ARPI, which evolved from the Group Risk Plan (GRP) introduced in 1993, focuses on yield insurance based on broader area outcomes rather than individual farm performances, thus offering a more efficient and cost-effective index insurance solution for risk management.<sup>2</sup> Although facing significant competition from farm-based insurance, recent data show that these area-based index insurance programs have accumulated liabilities exceeding \$9 billion, accounting for about 7% of the FCIP's total agricultural liability coverage, a substantial part of the total insurance coverage.<sup>3</sup>

Apart from the U.S., area-based insurance schemes have achieved greater success in developing countries. In 2016, the Indian government launched the nationwide area-yield crop insurance scheme, Pradhan Mantri Fasal Bima Yojana (PMFBY). Over the past eight years,

---

<sup>1</sup>For example, the USDA subsidizes premium costs for most acreage policies at the rate of about 60%. In addition, the government pays 100% of the premium for CAT policies (Rosch, 2021).

<sup>2</sup>The ARPI, operational since 2014, supersedes both the GRP and the Group Risk Income Protection Plan (GRIP).

<sup>3</sup>Data on area-yield insurance premiums is sourced from the USDA National Summary of Business, accessible at <https://www.rma.usda.gov/en/Information-Tools/Summary-of-Business>.

568 million farmers have enrolled in the program, with more than 262 million receiving claims totaling over \$20 billion. Enrollment continues to grow at a rate of 27% for FY2023-24<sup>4</sup>. Moreover, a comprehensive survey conducted by Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) covered 265 million agricultural policies written in 2020 across low and middle-income countries (LMIC)<sup>5</sup> (Hazell et al., 2021), indicating that approximately half of the farms operating in LMICs had some form of insurance coverage<sup>6</sup>. According to the survey, about 83% of agricultural insurance programs in LMICs are index-based, with around 32% of all index-based programs being area-yield index schemes, though this varies by region (50% in Latin America and the Caribbean, 37% in Africa, and 18% in Asia).

Area-yield insurance indemnifies policyholders based on the aggregate yield of a surrounding geographic area (i.e., a county) rather than individual farm-loss history, therefore, significantly reduces the concerns related to moral hazard and adverse selection, which are more prevalent in individual-yield insurance programs (Skees et al., 1997; Mahul, 1999). As such, area-yield insurance programs incentivize farmers' sustainable practices and investments such as implementing water conservation measures, diversifying their crops, adopting climate-smart agriculture techniques, etc. By aligning insurance incentives with environmental stewardship, area-yield insurance contributes to the promotion of sustainable agriculture.

Beyond its sustainability incentive, area-yield insurance presents notable operational efficiency. Since claims are triggered based on aggregate yields, area-yield contracts reduce insurers' administrative costs as less manpower is required for activities including record keeping, individual yield verification, and loss adjustment. Moreover, information about area yield is public and transparent. This means insurers are equally well informed about the distribution of area yield as individual producers are, and hence, the adverse selection problem is reduced (Miranda, 1991). Finally, the data limitation issue is largely addressed by area-yield insurance, because data are often more substantial and of higher quality at

---

<sup>4</sup>Statistics retrieved from the India Ministry of Agriculture & Farmers Welfare Press Information Bureau: <https://pib.gov.in/PressReleasePage.aspx?PRID=2011791>. According to this press release, the total claims paid to enrolled farmers since the start of this program amount to 1,560 billion Indian Rupees. The \$20 billion figure is calculated using the average Indian Rupee to U.S. dollar exchange rate of 0.013 over this period.

<sup>5</sup>See the definition of LMIC in (World Bank, 2023).

<sup>6</sup>Lowder et al. (2016) estimates that there are 515 million farms in the LMICs.

aggregate level than at individual farm level (Porth et al., 2019). These make index-based area-yield insurance particularly attractive to small-scale farmers who often lack access to formal insurance markets due to high costs, insufficient data, and complex administrative processes. Therefore, area-yield insurance has been shown to be a commercially viable and financially sustainable alternative to individual-based insurance (Barnett et al., 2005; Deng et al., 2007), and can substantially improve the agricultural credit market in the low-wealth region (Carter et al., 2007).<sup>7</sup>

While area-yield insurance programs show promise, one important assumption for their effectiveness in mitigating moral hazard is the size of the risk pool or the scale of individual farmers' operations (Bourgeon and Chambers, 2003). In practice, these insurance contracts are designed to cover sufficiently large geographical areas, ensuring that the impact of any single farmer on the area's overall yield is negligible. However, too large area makes the resulting contracts suffer from *basis risk*, referring to the potential mismatch between area-level outcomes and individual farm performances. As a result, there is trade-off between moral hazard and basis risk in designing area-yield insurance. The prevailing U.S. federal area-yield insurance program selects risk pools using county boundaries. However, risk pooling on administrative boundaries is rather arbitrary, as Skees et al. (1997) pointed out: "An area-index based on county yields is not ideal since county boundaries do not necessarily group together producers with similar patterns of year-to-year percentage deviations from forecasted yields."

This chapter aims to propose a behavior-based machine learning approach for optimal risk pooling in area-yield insurance design, by utilizing the moral hazard and basis risk trade-off. First, we determine the optimal number of risk pools,  $K$ , by analyzing producers' behavior with the existence of area-yield insurance. Area-yield insurance's virtue of being moral hazard proof relies on the assumption that no individual producer can influence the insurance payoff. Therefore, we select the largest  $K$  possible to minimize basis risk while still

---

<sup>7</sup>The success of weather index crop insurance programs in India serves as a prime example of its effectiveness. Weather index insurance was introduced to Indian farmers in 2003 and officially adopted by the national government in 2007 as an alternative to crop-yield index insurance. By 2012, it had expanded to cover 12 million farmers across 15 million hectares, safeguarding 40 different crops against adverse weather conditions. Source: Climate Change, Agriculture and Food Security (CCAFS), available at <https://ccafs.cgiar.org/outcomes/improved-rainfall-thresholds-index-insurance-india>.

keeping the majority of producers motivated with proper production activities, via achieving the insurer’s targeted compliance rate. With the determined optimal  $K$ , in the second step, we group producers into the  $K$  risk pools using unsupervised machine learning technique, i.e., spectral clustering method. Since greater risk reduction will be achieved if the individual yield is more highly correlated to the area yield (Miranda, 1991), it is intuitive to pool producers with similar production history together. Spectral clustering analysis is effective to address the challenges of high-dimensionality and computational complexity.

To empirically validate our sustainable risk pooling framework, an ideal comparison would involve contrasting farms pooled by our method against those grouped by conventional area-based plans, which typically use arbitrary county boundaries for risk pooling. However, due to the unavailability of detailed farm-level data across large regions, this direct comparison is impractical. To navigate this limitation, in the main results of this chapter, we apply our risk pooling methodology to the county-level corn production data in the U.S. Heartland Region. The findings from our analysis clearly indicate that the method we propose substantially reduces contract basis risk and effectively mitigates **tail risk** for producers. Benchmarking against risk pooling based on Agricultural Districts (AGD) boundaries, our proposed sustainable risk pooling method results in a basis risk reduction exceeding 29%. Compared with the state boundaries, the basis risk reduction is more than 55%. These improvements are also economically significant. For example, the enhancement on tail risk implies a yield improvement ranging from up to 12 bu/acre, equivalent to \$2.8 billion improvements on profits for the entire Heartland area. To further address concerns that county-level risk pooling may not accurately reflect farm-level risks, we reassess the performance of our framework using a simulated farm-level dataset. This simulation serves as a proxy to reassure the effectiveness of our proposed framework under more granular, farm-specific conditions.

We examine two methods to understand the mechanism behind the observed superior performance. First, we apply our proposed framework to data from Illinois, given its distinctive terrain. This application allows us to discern the geographic implications of risk pooling. Our results indicate that the terrain characteristics of Illinois are reflected in the risk pooling outcomes, lending geographic support to our method. Second, we contrast our

behavior-based method with the well-established gap statistic method to determine the optimal  $K^*$ . This comparison highlights that while the gap statistic method is valuable, it lacks the necessary economic context and fails to consistently identify an effective  $K^*$  for reducing basis risk.

We further verify the robustness of the proposed risk pooling procedure in several ways. First, we discuss impact of an insurer’s targeted sustainable compliance ratio on the basis risk improvements. Second, we test the impact of high-dimensionality for area-yield insurance design. Third, we evaluate the robustness of performance against varying cost assumptions. Fourth, we consider alternative risk measures in assessing the risk pooling performance. Fifth, we allow producers the flexibility to adjust their coverage ratios according to their individual preferences. Finally, we impose the same-size constraint in the risk pools to avoid the formation of very small risk pools.

This project contributes to the area-yield insurance literature. In his seminal paper, [Miranda \(1991\)](#) discusses the effectiveness of an area-yield insurance program from the variance reduction perspective. A similar setting is also considered by [Smith et al. \(1994\)](#). [Mahul \(1999\)](#) and [Ramaswami and Roe \(2004\)](#) solve for an optimal area-yield insurance contract design in a utility maximization framework. The literature also discusses producers’ behaviors and decisions. For example, [Chambers and Quiggin \(2002\)](#) study the optimal level of enrollment and the production decisions in a joint decision problem. [Bourgeon and Chambers \(2003\)](#) design optimal area-yield insurance in the presence of asymmetric information. Beyond its benefit in addressing moral hazard, more recently, the literature highlights the benefit of cost-effectiveness of area-yield insurance, making it less reliant on government subsidies. This makes area-yield insurance a more affordable option in both developed and developing markets ([Gong et al., 2023](#); [Ye et al., 2020](#); [Kusumaningrum et al., 2021](#)). Moreover, many studies develop better crop yield models by incorporating spatial and temporal dependence, which will help improve area-yield insurance pricing (see, e.g., [Wang and Zhang, 2003](#); [Harri et al., 2011a](#); [Annan et al., 2014](#); [Liu and Ker, 2021](#)). [Sethanand et al. \(2023\)](#) implemented a machine learning algorithm, specifically Random Forest, to enhance the accuracy of yield predictions and improve the adoption of crop insurance schemes through tech-

nological advancements. This project studies the optimal risk pooling in designing area-yield insurance through an unsupervised learning method, and incorporates producers' behavior when optimally select the number of risk pools,  $K$ .

This study contributes to the fields of behavioral operational research (BOR) and behavioral economics. Behavioral analytic frameworks have been widely applied in portfolio selection (Barberis et al., 2001; Barberis and Huang, 2008; Fulga, 2016; Hwang et al., 2018; Harris and Mazibas, 2022) and insurance decision-making (Chambers and Quiggin, 2002; Barseghyan et al., 2013; Elabed and Carter, 2015; Clarke, 2016). Ackermann (2024) highlights the importance of integrating Problem Structuring Methods (PSM) with BOR to effectively address complex systemic issues, such as climate change. This study generally follows the analytic framework described by Mintz et al. (2023). By incorporating producers' behavioral insights into the optimization of area-yield insurance risk pooling, we address the decision-making processes of producers and insurers, thereby enhancing the effectiveness in the designing and adoption of area-yield insurance programs.

This project also broadly belongs to the literature of index-based securities. The demand and efficiencies of index securities have been studied by many researchers (Cole et al., 2014; Chantararat et al., 2017; Casaburi and Willis, 2018; Cai et al., 2020). Basis risk is commonly cited as one of the most challenging issues that causes their low demand (Clarke, 2016; Cummins et al., 2004), among others. Jensen et al. (2016) evaluate the index-based livestock insurance (IBLI) product in northern Kenya. They find that while the policy reduces covariate risk and downside risk substantially, policyholders are left with large basis risk. Chantararat et al. (2013) proposes a new IBLI based on predicted area-average livestock mortality to reduce basis risk. Chen et al. (2023a) design a novel weather index insurance that can effectively reduce basis risk by embedding a neural network-based expected utility maximization scheme. In this study, we focus on reducing the basis risk of area-yield insurance, and propose to improve the insurance performance from a sustainable risk pooling perspective.

The remainder of this chapter proceeds as follows. Section 2.2 introduces the area-yield program and sets up the problem. In section 2.3, we introduce the proposed behavior-

based machine learning framework. Section 2.4 presents the empirical implementation of our model. Section 2.5 verifies the robustness of the proposed risk pooling procedure. Section 2.6 provides further validation by testing the model’s performance on a simulated farm-level production dataset. Section 2.7 concludes. Details of proofs and additional empirical analysis are collected in Appendices.

## 2.2 Area-yield Insurance Program

The area-yield insurance program was first introduced in 1993, initially targeting soybean farmers in selected U.S. counties. Over time, the program expanded its coverage to include major commodities such as corn, wheat, and cotton (GIIF, 2024). A key feature of area-yield insurance policies is their method of indemnity payment, which is based on the average yield within a county rather than on the loss of individual farms. This approach effectively mitigates issues related to information asymmetry, including moral hazard and adverse selection, as the actions of any single producer are unlikely to have a substantial impact on the overall yields at the county level.

Area-yield insurance organizes a group of producers into  $K$  distinct risk pools. Consider producer  $i$  operating within the  $k$ th risk pool. Under the protection of area-yield insurance, this producer is eligible to receive an indemnity payment if the area yield of the  $k$ th risk pool, denoted as  $y_k$ , falls below a pre-established threshold,  $\bar{y}_c$ . The indemnity function, applicable universally within this pool, is formulated as follows:

$$I_k = \max(\bar{y}_c - y_k, 0) \times scale, \tag{2.1}$$

where  $\bar{y}_c$  is the critical yield, calculated as

$$\bar{y}_c = \mu_k \times coverage, \tag{2.2}$$

and  $\mu_k$  represents the expected area yield level, that is,  $\mu_k = \mathbf{E}[y_k]$ . *coverage* and *scale* offer producers additional flexibility. For example, in the US area-yield insurance program,

producers can choose a *coverage* level that ranges from 70% to 90%, and a *scale* level from 90% to 150%. Typically farm-level volatility is higher than the county-level. Increasing their *scale* and *coverage* ensures that producers can have sufficient coverage in years when production significantly drops.

Miranda (1991) establishes the classic connection between individual yield  $y_i$  and area yield  $y_k$ , a relationship that is widely accepted in the area-yield insurance literature (Skees et al., 1997; Mahul, 1999; Barnett et al., 2005; Harri et al., 2011b). By projecting the producer’s individual yield  $y_i$  onto the area yield  $y_k$ ,  $y_i$  can be modeled as follows:

$$y_i = \mu_i + \beta_i \cdot (y_k - \mu_k) + \epsilon_i, \quad (2.3)$$

where

$$\beta_i = \text{Cov}(y_i, y_k) / \sigma_{y_k}^2 \quad (2.4)$$

$$\text{E}[\epsilon_i] = 0 \quad \text{Var}[\epsilon_i] = \sigma_{\epsilon_i}^2 \quad \text{Cov}(y_k, \epsilon_i) = 0 \quad (2.5)$$

$$\text{E}[y_i] = \mu_i \quad \text{Var}[y_i] = \sigma_{y_i}^2 \quad (2.6)$$

$$\text{E}[y_k] = \mu_k \quad \text{Var}[y_k] = \sigma_{y_k}^2. \quad (2.7)$$

Here, Equation (2.3) decomposes individual yield variation into a systemic component  $\beta_i \cdot (y_k - \mu_k)$  that is perfectly correlated with the area yield and a non-systemic component  $\epsilon_i$  that is uncorrelated with area yield. The coefficient  $\beta_i$  quantifies the sensitivity of producer’s individual yield to the systemic factors that affect the area yield.<sup>8</sup>

We note that Miranda’s model is limited in capturing tail dependence. To partially address this concern, our empirical analysis in Section 2.4 uses a relatively large time window to better reflect the less frequent extreme events reflected into the yield dependence structure among producers. However, due to its simplicity and wide acceptance, we use this model as assumptions of Proposition 2.1 and Proposition 2.2 for better illustration and tractability. Our propositions should remain valid if the model is adjusted to better account for tail risk,

---

<sup>8</sup>Given that the area-yield insurance contract aims to hedge the systemic risk faced by producers,  $\beta_i$  needs to be positive to ensure the contract’s effectiveness.

as they are supported by our empirical analyses.

## 2.3 Sustainable Risk Pooling Framework

Typically area-yield index-based insurance programs, such as the Area Risk Protection Insurance (ARPI) in the US, defines risk pools based on county boundaries. While this method provides a structured approach to risk assessment, relying on county yields for an area-index may not be optimal, as county boundaries often fail to accurately group together producers with similar year-to-year percentage deviations from forecast yields. This mismatch can lead to a misalignment between the actual risk profiles of individual producers and the broader risk pool defined by these administrative boundaries (Skees et al., 1997).

To mitigate basis risk, producers protected by area-yield insurance seek to minimize their self-insured idiosyncratic component,  $\epsilon_i$ . Thus, this chapter aims to propose a sustainable risk pooling framework that incorporates producers' behaviour and employs a data-driven approach for an optimal area-yield insurance design. This framework comprises two essential steps:

1. **Determining the optimal number of risk pools:** This step strikes the best balance between reducing basis risk and managing moral hazard.
2. **Grouping producers with similar risks:** This steps tries to improve the effectiveness of the designed area-yield index insurance by creating homogeneous risk groups.

This two-step process ensures that the systemic part of producers' risks are appropriately hedged, while the impact of idiosyncratic part is minimized, contributing to a more robust and resilient area-yield index insurance mechanism.

### 2.3.1 Determining the Optimal Number of Clusters: A Behavior-based Approach

Selecting the optimal number of clusters, denoted as  $K^*$ , is a critical step in the efficient risk-pooling procedure for area-yield insurance design. As mentioned earlier, the determination of  $K^*$  strikes a crucial trade-off between moral hazard and basis risk. More specifically,

- **A larger  $K$ :** Increasing the number of risk pools, denoted as  $K$ , leads to a reduction in basis risk. This is because a larger number of pools allows more accurate reflection of individual producers' losses and a fairer premium collection mechanism. However, this benefit is accompanied by an elevated risk of moral hazard. In smaller pools, each producer has greater influence over the yield of their respective pools, potentially leading to higher adverse incentives.
- **A smaller  $K$ :** Conversely, a smaller  $K$  mitigates the moral hazard issue, as individual producers possess less influence over the collective area yield. This reduction in moral hazard comes with a cost of higher basis risk. That is, the insurance contract may become less sensitive to the specific losses of individual producers due to the broader nature of the risk pools, thus increasing basis risk.

Driven by this trade-off between moral hazard and basis risk, this chapter introduces a behavior-based methodology to determine the optimal number of risk pools,  $K^*$ . Our objective is to identify the maximum feasible  $K$  that effectively minimizes basis risk, while concurrently maintaining high levels of engagement and motivation among the majority of producers. In doing so, we adopt a behavioral analytics framework as follows. First, we develop an expected utility maximization model to analyze producers' agricultural practice decisions under the protection of area-yield insurance contracts. Next, we use historical yield data to estimate the distribution for each producer and predict their future decisions within the risk pooling setting of each  $K$ . Finally, we optimize the value of  $K$  based on these predictions to design risk-pooling incentives that ensure the majority of producers adhere to moral-hazard-free practices. Such framework is generally consistent with the analytic framework described by [Mintz et al. \(2023\)](#).

Let us consider a producer, denoted as  $i$ , who is assigned to the  $k$ th risk pool and possesses a non-pledgeable initial endowment,  $w_0$ . The indemnity pay-off of this producer,  $I_k$ , is defined in Equation (2.1). Let  $\mathbf{c}_i = (c_{i1}, \dots, c_{iJ})$  denote the cost vector of her inputs, where each component  $c_{ij}, j = 1, 2, \dots, J$  corresponds to a specific input cost, including the insurance premium payment. With the protection of area-yield insurance, the terminal wealth of producer  $i$  is  $w_0 + y_i - \sum_{j=1}^J c_{ij} + I_k$ . In this chapter, we focus on optimizing the area-yield insurance for production loss, without taking into account crop price risk, but the proposed framework can easily be extended to include protection against price risk.

Given these conditions, the objective of producer  $i$  is to maximize her expected utility through strategically managing her inputs  $\mathbf{c}_i$ . Specifically, let  $u(\cdot)$  be a von Neumann-Morgenstern utility function of wealth for producer  $i$ , her objective function is:

$$\max_{\mathbf{c}_i \in \mathcal{C}} \mathbf{E} \left[ u \left( w_0 + y_i - \sum_{j=1}^J c_{ij} + I_k \right) \right], \quad (2.8)$$

where  $\mathcal{C} \subseteq \mathbb{R}_+^J$  defines the feasible set of the input cost vector  $\mathbf{c}_i$ . Hereafter, we denote the total cost of production as  $C_i = \sum_{j=1}^J c_{ij}$ , and her expected utility as  $EU(\mathbf{c}_i) = \mathbf{E}[u(w_0 + y_i - C_i + I_k)]$ .

To simplify the analysis, we assume that a producer can choose between two different farming strategies: active production management (Strategy  $A$ ) and passive production management (Strategy  $B$ ). These two strategies are associated with different yield distributions,  $y_i^A$  and  $y_i^B$ , and distinct input cost vectors,  $\mathbf{c}_i^A$  and  $\mathbf{c}_i^B$ . Producer  $i$  selects the optimal strategy,  $\mathbf{c}_i^*$ , where  $\mathbf{c}_i^* \in \{\mathbf{c}_i^A, \mathbf{c}_i^B\}$ , to maximize her expected utility. Intuitively, Strategy  $A$  involves a higher total cost compared to Strategy  $B$  (i.e.,  $C_i^A \geq C_i^B$ ), but has a superior yield distribution, i.e.,  $y_i^A$  has first-order stochastic dominance over  $y_i^B$ .<sup>9</sup>

We use an indicator function to represent producer  $i$ 's optimal choice,  $\mathbf{c}_i^*$ , is an active

---

<sup>9</sup>In our context, the yield  $y_i^A$  from producer  $i$  having first order stochastically dominance over  $y_i^B$  is equivalent to  $\mathbf{P}(y_i^A \geq y) \geq \mathbf{P}(y_i^B \geq y)$  for all possible yields in the range  $y_{min} \leq y \leq y_{max}$ . Additionally, for some  $y_{min} \leq y \leq y_{max}$ ,  $\mathbf{P}(y_i^A \geq y) > \mathbf{P}(y_i^B \geq y)$ .

management strategy, i.e.,

$$\mathbf{1}(\mathbf{c}_i^*) = \begin{cases} 1 & \text{if } \mathbf{c}_i^* = \mathbf{c}_i^A, \\ 0 & \text{otherwise.} \end{cases} \quad (2.9)$$

A rational producer would choose  $\mathbf{c}_i^* = \mathbf{c}_i^A$  if and only if  $EU(\mathbf{c}_i^A) \geq EU(\mathbf{c}_i^B)$ . Notably, producers' choices of strategies are affected by both the risk pooling outcome and their own characteristics, as described in the two propositions below.

**Proposition 2.1.** *Consider two producers, labeled as  $i$  and  $j$ , both operating within the  $k$ th risk pool. Let  $\beta_i^A$  and  $\beta_j^A$  be the sensitivities of their respective yields to systemic factors under active management, and  $\sigma_{\epsilon_i^A}^2$  and  $\sigma_{\epsilon_j^A}^2$  be the variances of the non-systemic components of their yields under active management, as defined in Equations (2.4) and (2.5), respectively. Ceteris paribus, we have the following:*

- (a) *The producer more responsive to systemic factor under active management is more likely to take active production strategy. That is,  $\mathbf{1}(\mathbf{c}_i^*) \geq \mathbf{1}(\mathbf{c}_j^*)$  if  $\beta_i^A \geq \beta_j^A$ .*
- (b) *The producer with lower idiosyncratic variance under active management is more likely to take active production strategy. That is,  $\mathbf{1}(\mathbf{c}_i^*) \geq \mathbf{1}(\mathbf{c}_j^*)$  if  $\sigma_{\epsilon_i^A}^2 \leq \sigma_{\epsilon_j^A}^2$ .*

Proposition 2.1 implies that producers with yields highly responsive to the area yield of their respective risk pool and experiencing minimal self-insured idiosyncratic risks are more inclined to implement active farm management practices. This proposition indicates that it is important to group producers with similar risks into the same risk pools for optimal area-yield insurance risk pooling.

**Proposition 2.2.** *Consider two producers, denoted as  $i$  and  $j$ , both operating within the  $k$ th risk pool. Let  $E[y_i^s]$  and  $E[y_j^s]$  be their expected yield through production management strategy  $s \in \{A, B\}$ . Let  $C_i^s$  and  $C_j^s$  be their total cost under production management strategy  $s \in \{A, B\}$ . Ceteris paribus, we have the following:*

- (a) *If  $E[y_i^A] \geq E[y_j^A]$ , then  $\mathbf{1}(\mathbf{c}_i^*) \geq \mathbf{1}(\mathbf{c}_j^*)$ .*

(b) If  $E[y_i^B] \leq E[y_j^B]$ , then  $\mathbf{1}(\mathbf{c}_i^*) \geq \mathbf{1}(\mathbf{c}_j^*)$ .

(c) If  $C_i^A \leq C_j^A$ , then  $\mathbf{1}(\mathbf{c}_i^*) \geq \mathbf{1}(\mathbf{c}_j^*)$ .

(d) If  $C_i^B \geq C_j^B$ , then  $\mathbf{1}(\mathbf{c}_i^*) \geq \mathbf{1}(\mathbf{c}_j^*)$ .

Proposition 2.2 indicates that producers with higher (lower) active (passive) yields, and producers with lower (higher) active (passive) costs, are more inclined towards committing to active production. In essence, those who stand to benefit significantly from transitioning to active management are more likely to make a commitment to active management. Detailed proofs of both Proposition 2.1 and Proposition 2.2 are collected in Appendix 2A.1.

To evaluate the effectiveness of an area-yield insurance program, it is essential to conduct an aggregate analysis of producers' farming behavior. To facilitate this analysis, we introduce the concept of Attained Compliance Ratio (ACR), which quantifies the expected proportion of producers who choose active production methods. For a given value of  $K$ , where producers are assigned optimally, the ACR is given by

$$\text{ACR}(K) = \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N \mathbf{1}(\mathbf{c}_i^*(K)) \right], \quad (2.10)$$

where  $N$  is the total number of producers.  $\mathbf{1}(\mathbf{c}_i^*(K))$  symbolizes the behavior of producer  $i$  in relation to the number of risk pools  $K$ . As established, producers' optimal behavior changes in response to variations in  $K$ .

From the insurer's perspective, the area-yield insurance program is effective when a majority of the enrolled producers maximize their utility with active farm management. This ensures that the moral hazard inherent to the insurer's portfolio remains tolerably low. If the insurer sets a Sustainable Compliance Ratio (SCR), which represents the proportion of producers engaging in active management required by the insurer to maintain a sustainable area-yield insurance program, the optimal  $K^*$  is determined by ensuring that the ACR derived from the risk pooling outcomes aligns with SCR. To elaborate:

$$K^* = \underset{K}{\operatorname{argmin}} |\text{ACR}(K) - \text{SCR}|, \quad (2.11)$$

### 2.3.2 Optimal Risk Pooling Algorithm

Proposition 2.1 implies that, for a given value of  $K$ , it is important to group producers with similar risk profiles into the same risk pools. The proposed behavioral risk pooling framework can accommodate any dependence structure that informs the relationship between pairs of producers. In this project, we use the empirical correlation matrix to serve this purpose, thereby avoiding the concern of model uncertainty introduced by complex dependence structures. Consequently, we can employ a clustering algorithm, effectively grouping producers based on comparable risk levels.

However, traditional clustering approaches use yield correlation matrices, which encounters the challenge of high dimensionality. Particularly when dealing with a large number of producers, the high-dimensional data not only requires significant computational resources in terms of time and memory but may also impair the effectiveness of conventional distance-based clustering algorithms.

To circumvent these challenges, we introduce the spectral clustering method into the risk pooling procedure, which first reduces the dimensionality of the correlation matrix through spectral decomposition and then utilizes standard distance-based clustering techniques on the transformed data with a lower dimension. Let  $M = (\rho_{ij})_{N \times N}$  denote the correlation matrix of historical yields, where  $N$  represents the total number of producers. The spectral decomposition of  $M$  is expressed as  $M = QAQ^{-1}$ , with  $A$  being a diagonal matrix comprising the eigenvalues of  $M$ , arranged in descending order. The columns of  $Q$  are the corresponding eigenvectors of  $M$ . We then construct a new  $N \times d$  matrix,  $Z$ , by selecting the top  $d$  eigenvectors, represented as  $Z = [Q_{:1}, \dots, Q_{:d}]$ , to form distinct clusters. A succinct summary of the spectral clustering method is presented in Algorithm 1.

Compared with the traditional distance-based K-means algorithm, the spectral clustering method offers several distinct advantages. First, it effectively handles the challenges posed by high-dimensional data. The phenomenon known as the “curse of dimensionality” challenges distance-based clustering algorithms when dealing with high-dimensional data (Beyer et al., 1999). Second, spectral clustering presents efficiency in both time and memory usage.

---

**Algorithm 1:** Risk pooling based on spectral clustering

---

**Input:**

- Squared data  $M = (\rho_{ij})_{N \times N}$  (correlation matrix of the yield data).
- Number of risk pools  $K$ .
- Desired proportion of variance retained  $q$ .

**Steps:**

1. **Spectral Decomposition of  $M$ :** Decompose  $M$  as  $M = QAQ^{-1}$ . Here  $A$  is a diagonal matrix containing the eigenvalues of  $M$ , while the columns of  $Q$  represent the eigenvectors of  $M$ .
2. **Eigenvalue Ordering:** Sort the eigenvalues in descending order:  $A_{11} > A_{22} > \dots > A_{NN}$ . Arrange the eigenvectors in corresponding order.
3. **Eigenvector Selection:** Choose the first  $d$  eigenvalues such that their sum as a fraction of the sum of all eigenvalues is no less than  $q$ , that is,

$$\frac{\sum_{i=1}^d A_{ii}}{\sum_{i=1}^N A_{ii}} \geq q.$$

4. **Matrix Formation:** Construct a  $N \times d$  matrix,  $Z$ , using the first  $d$  eigenvectors:

$$Z = [Q_{:1}, \dots, Q_{:d}].$$

5. **Risk-Pooling via K-means:** Apply the K-means clustering algorithm to the dataset  $Z$  to form the risk pools.

**Output:**

- The result of risk-pooling, denoted as a  $1 \times N$  vector  $(k_1, \dots, k_N)$ , where  $k_i$  (for  $i = 1, \dots, N$ ) is an element from the set  $\{1, \dots, K\}$ . This element represents the assignment of producer  $i$  to one of the  $K$  risk pools.
-

## 2.4 Empirical Analysis

In this section, we implement the proposed sustainable risk pooling framework on corn production data from the U.S. Heartland Region. Section 2.4.1 introduces the data used in our empirical analysis and describes pre-analysis process. Section 2.4.2 presents our risk pooling results. In Section 2.4.3, we conduct a subsample analysis in the state of Illinois and focus on economic meaning of the risk pooling. Finally, in Section 2.4.4, we compare our proposed method for determining the optimal  $K$  with a widely accepted statistical method, the gap statistic.

### 2.4.1 Data

#### Production data

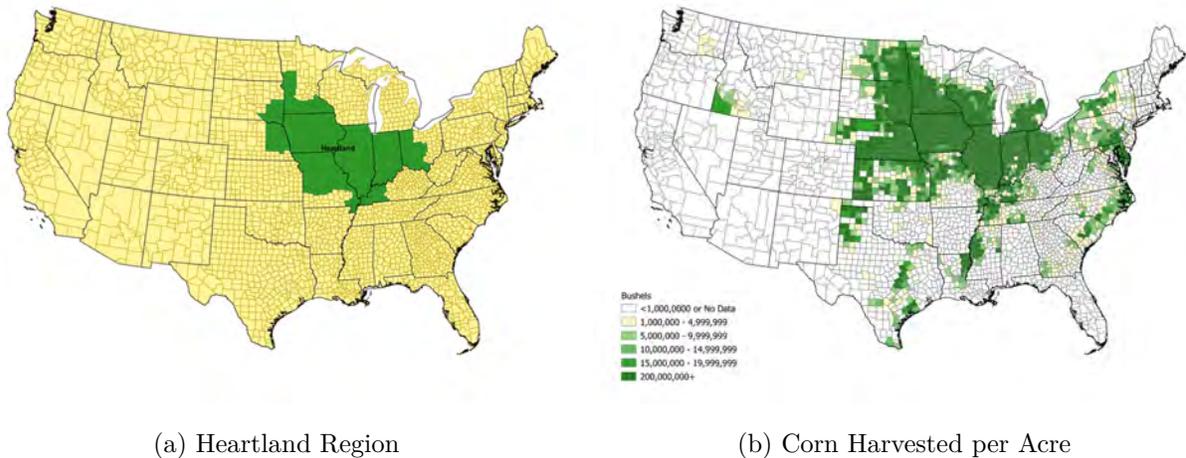
Due to the unavailability of farm-level production data, we have resorted to the most detailed alternative: an annual county-level corn production dataset from the National Agricultural Statistics Service (NASS).<sup>10</sup> Our study focuses on the counties within the Heartland Region, a key area for corn production, as depicted in Figure 2.1(a). The data is representative because corn is the most valuable agricultural commodity in the US, and the Heartland Region includes major corn-producing states such as Iowa, Illinois, Indiana, etc. Moreover, it encompasses 543 counties and leads among the nine U.S. farm resource regions with the highest number of farms (22%), the largest cropland area (27%), and the greatest production value (23%).<sup>11</sup> Our analysis reveals that from 2015 to 2019, these Heartland counties contributed to 65% of the total U.S. corn production (see Figure 2.1(b)).

The sample period spans from 1940 to 2019, and the summary statistics of the data are provided in Table 2.1. To account for technological advancements and other long-term yield influences, following (Deng et al., 2007), a detrending process using a second-order

---

<sup>10</sup>To address concerns that using county-level data as a proxy for farm-level data might lead to misinterpretation of risks, we conduct farm-level analysis based on simulation in Section 2.6, as a robustness check for our proposed sustainable risk pooling method.

<sup>11</sup>Source: USDA National Agricultural Statistics Service, Economic Research Service.



**Figure 2.1.** Heartland region and U.S. corn production (Illustrative period of 2015-19).

polynomial is applied to the raw production data to align the yield data with 2019 levels. Approximately 2.0% of the county-year data points are missing in the initial dataset, which we address through linear interpolations. Consequently, our processed dataset includes 43,440 county-year observations, encompassing 543 counties over 80 years<sup>12</sup>. For the purposes of clustering, we normalize both yield and cropland size data using the min-max scaler method.

## Cost data

To determine the cost of production, we reference the Heartland Region’s corn production cost data published by the USDA. Since 1996, the USDA has conducted five comprehensive censuses of corn production costs in the years 1996, 2001, 2005, 2010, and 2016.<sup>13</sup> However, there are notable discrepancies between this cost data and our yield data, both in terms of the timespan covered and the level of details. Due to these constraints, we approximate

<sup>12</sup>As established, to partially address the concern that Equation (2.3) is weak in capturing tail dependence, and given that we use a correlation matrix to capture pairwise relationships among producers—which is also not specifically tailored for tail risk—we utilize a relatively long window of historical yield data. This approach allows the model to capture some less frequent extreme events. It is important to note that such extensive historical yield data is not mandated by the proposed behavioral risk pooling framework if reliable tail dependence information is available.

<sup>13</sup>The full report can be accessed at <https://www.ers.usda.gov/data-products/commodity-costs-and-returns/>.

**TABLE 2.1**  
**Descriptive Statistics**

Our sample of 543 counties located in 9 states of the Heartland region of the U.S. corresponds to 43,440 county-year observations (Source: NASS). 43 million acres of corn production cropland are included. Panel A summarizes the corn yield data, and Panel B summarizes the corn cropland size data.

<i>Panel A. Corn Yield (BU/Acre) – Summary Statistics</i>							
State	Mean	Standard Deviation	Skewness	Excess Kurtosis	1st Quartile	Median	3rd Quartile
Illinois	194	31	-0.34	-0.41	165	189	209
Indiana	172	24	-0.38	-0.34	150	170	185
Iowa	213	26	-0.19	0.16	190	210	223
Kentucky	188	25	-0.14	-1.07	161	187	203
Minnesota	185	22	-0.84	0.35	173	190	201
Missouri	157	26	0.47	0.26	131	147	167
Nebraska	210	20	-0.43	-0.32	195	208	222
Ohio	169	22	-0.52	0.03	158	170	184
South Dakota	172	28	-0.28	-0.75	150	173	192

<i>Panel B. Corn Cropland Size (Acre) – Summary Statistics</i>							
State	Mean	Standard Deviation	Skewness	Excess Kurtosis	1st Quartile	Median	3rd Quartile
Illinois	99233	67991	1.11	1.06	49962	99233	133526
Indiana	56156	29267	0.14	-0.58	34453	56156	74996
Iowa	116275	43832	0.20	0.23	87850	116275	147041
Kentucky	30472	17997	0.74	-0.71	17293	30472	36535
Minnesota	109425	42361	0.21	-1.10	74470	109425	148958
Missouri	32805	23722	1.38	2.48	15644	32805	41945
Nebraska	113756	35305	-0.11	-0.92	89191	113756	139319
Ohio	51276	24365	-0.24	-0.59	36976	51276	68927
South Dakota	95210	28316	0.13	-0.95	74279	95210	112653

producers' total costs per acre  $C_i$  in relation to their expected yields  $\mu_i$ , as follows:

$$\begin{cases} C_i^A = (C_{variable} + C_{fixed}) \times \mu_i, \\ C_i^B = C_{fixed} \times \mu_i. \end{cases} \quad (2.12)$$

Here,  $C_{variable}$  denotes the variable cost ratio and  $C_{fixed}$  represents the fixed cost ratio. Segmenting the cost into its variable and fixed components allows for a nuanced understanding of a producer's management choices. More specifically, when a producer selects for an active production management strategy (i.e.,  $\mathbf{c}_i^* = \mathbf{c}^A$ ), the incurred cost encompasses both the variable and fixed costs. In contrast, if a producer selects a passive approach to production management (i.e.,  $\mathbf{c}_i^* = \mathbf{c}^B$ ), only the fixed cost is considered.

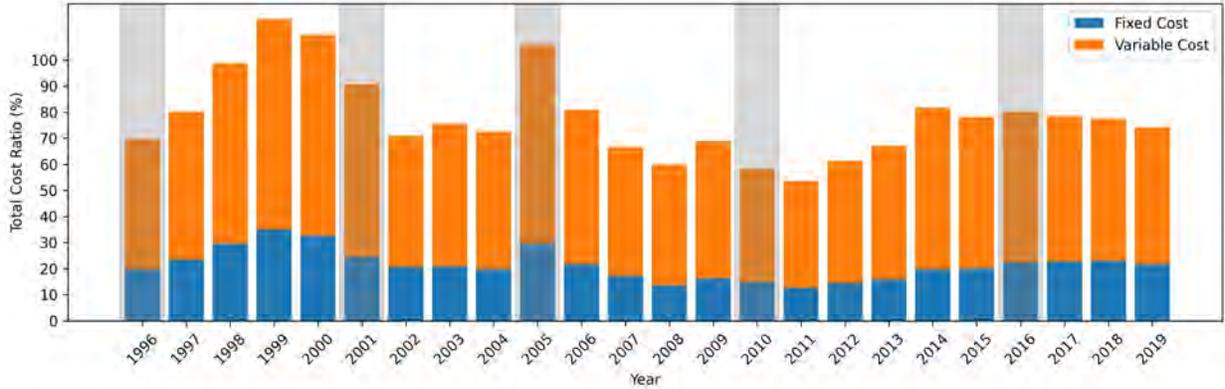
Producers' incentives should be determined ex-ante. This implies that producers would choose their strategies based on expected fixed and variable costs. Given that our dataset is detrended to account for inflation and technological changes, maintaining static expected ratios of fixed and variable costs is deemed acceptable. Moreover, using annual cost data might introduce a forward-looking bias in risk pooling analyses. Consequently, we compute the fixed and variable cost ratios for the Heartland Region for each year within the cost data's timeframe, using their average as the expected fixed and variable cost ratios for producers.<sup>14</sup> The fixed and variable cost ratios for the Heartland Region by year is presented in Figure 2.2. The grey bars highlights the census years. We check the robustness of the cost function assumptions in Section 2.5.3.

## 2.4.2 Baseline Results of the Sustainable Risk Pooling Framework

In our baseline analysis, we focus on an insurer whose objective is to motivate at least 95% of the enrolled producers to actively manage their farms to keep the moral hazard at a tolerable level. This is achieved by solving Equation (2.11), where we set the sustain-

---

<sup>14</sup>Here, the fixed cost ratio for each year is determined as the ratio of ownership cost (including capital recovery of machinery and equipment, cost of insurance, taxes, and general farm overhead) to total gross production value, and the variable cost ratio is computed as the ratio of operating cost (including cost of seed, fertilizer, chemicals, custom services, energy, repairs, purchased irrigation water, labor, interest on operating capital) to total gross production value.



**Figure 2.2.** The ratios of fixed and variable costs to value of corn produced for heartland region. The grey bars highlight the census years.  
*Source:* USDA Economic Research Service.

able compliance ratio (SCR) at 0.95. For additional robustness, alternative SCR levels are examined in Section 2.5.1.

We construct a yield correlation matrix for the producers and perform spectral decomposition to extract the principal components. For spectral clustering, we select the first 36 principal components, which account for 95% of the total variance. The robustness of the variance retained is tested in Section 2.5.2, where we find that our results remain consistent within a variance retention range of 85% to 99%.

### Solving for optimal production behaviors

To determine producers' production behavior (i.e., optimal choice of input strategy  $\mathbf{c}_i^*$ ), we solve Problem (2.8) for each individual producer's optimal decision,  $\mathbf{c}_i^*$ . In doing so, we adopt a model-free, non-parametric approach following [Chen et al. \(2023a\)](#). This approach involves substituting the theoretical variables in the expected utility function with their empirical equivalents, and thereby sidestepping the model risk associated with jointly modeling the high-dimensional distribution of individual yield  $y_i$  and the indemnity payment  $I_k$ . Specifically, for a random sample of  $(y_i, I_k) : \{(y_{it}, I_{kt})\}_{t=1, \dots, T}$ , problem (2.8) can be formulated as follows:

$$\begin{aligned}
\max_{\mathbf{c}_i \in \mathcal{C}} \mathbf{E} [U(\mathbf{c}_i)] &= \max_{\mathbf{c}_i \in \mathcal{C}} \mathbf{E} [u(w_0 + y_i - C_i + I_k)], \\
&= \max_{\mathbf{c}_i \in \mathcal{C}} \frac{1}{T} \sum_{t=1}^T u(w_0 + y_{it} - C_i + I_{kt}).
\end{aligned} \tag{2.13}$$

To ensure robustness, we examine a parametric approach, as detailed in Appendix 2A.2.

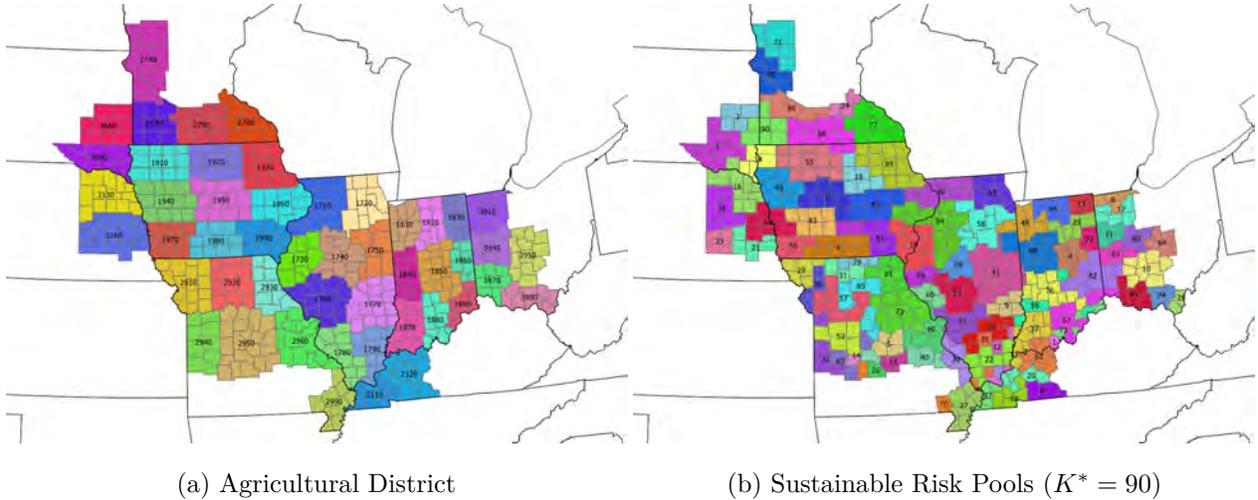
In the baseline analysis, we calculate  $I_k$  assuming producers uniformly set the *scale* and *coverage* of their policies at 100%. In the robustness checks in Section 2.5.5, we find that the effectiveness of the proposed sustainable risk pooling method remain unaffected when producers choose their optimal *scale* and *coverage* (Mahul, 1999; Vercaemmen, 2000).

For producers' risk preference, we consider both the Constant Absolute Risk Aversion (CARA) model and the Constant Relative Risk Aversion (CRRA) model. More specifically, we use an exponential utility function  $u(w) = -\frac{1}{A}e^{-Aw}$  and an absolute risk aversion parameter  $A = 0.5$  for CARA, and a power utility function  $u(w) = \frac{w^{1-\gamma}-1}{1-\gamma}$  with a relative risk aversion parameter  $\gamma = 2.0$ . In an untabulated analysis, we find that our results remain robust for various choices of risk aversion parameters for both utility models.

## Benchmarks and risk measures

As previously established, owing to the unavailability of actual farm-level data, we use county-level yield data as a proxy for farm-level data. This limitation prevents empirical evaluation against real county boundaries. Instead, we assess our proposed risk pooling method against two alternative arbitrary boundary sets: Agricultural Districts (AGD) and State boundaries. Agricultural Districts are defined by NASS, grouping counties within each state based on geographic, climatic, and agricultural characteristics. Figure 2.3(a) illustrates the AGDs in the Heartland Region. While these boundaries reflect certain geographical and environmental characteristics, their use in risk pooling is not specifically tailored to individual commodities.

To evaluate the effectiveness of risk pooling, we assess various basis risk and tail risk measures, including [Mean Squared Error](#) (MSE) for basis risk reduction, [Value-at-Risk](#) (VaR)



**Figure 2.3.** Agricultural District of the heartland region and the proposed sustainable risk pooling results.

and [Expected Shortfall](#) (ES) on profits for downside risk reduction. Other alternative measures for assessing risk pooling quality are further explored in [Section 2.5.4](#).

### Performances of the sustainable risk pooling framework

We evaluate the effectiveness of area-yield insurance derived from risk pools generated through our sustainable risk pooling method in [Table 2.2](#). This table summarizes the improvements of sustainable risk pools compared to those based on agricultural districts and state boundaries. These improvements are quantified by basis risk (MSE) and tail risk (VaR and ES). We report results with risk measures aggregated through both equal weighting and cropland-size weighting approaches.

Our optimal risk pooling method suggests an optimal  $K^*$  of 94 under the exponential utility and 92 under the power utility. This indicates that our results are not sensitive to utility assumptions. The improvements in MSE versus AGD range from 21.7% to 30.2%, and range from 53.0% to 58.9% compared to state boundaries, across different risk aversion assumptions and risk aggregation methods.

Panel B and Panel C report the improvements in tail risk protection. Our proposed sustainable risk pooling method enhances the 5%-level VaR on profits by 1.7% to 8.0% and the 5%-level ES by 2.3% to 8.4%, across different risk aversion assumptions and risk

aggregation methods. These improvements are economically significant. In particular, the enhancement on VaR implies a yield improvement ranging from 2.55 bu/acre to 12 bu/acre, equivalent to \$598 million to \$2.8 billion improvements on profits for the Heartland area, after considering the corn price and land size of the area. Similarly, the improvements in ES correspond to 3.22 bushes/acre to 11.76 bushes/acre of yield improvements, translating to profit enhancement ranging from \$755 million to \$2.76 billion for the Heartland area.<sup>15</sup>

In summary, results validate the effectiveness of sustainable risk pooling approach in optimizing area-yield insurance for producers. Since the positive risk improvements are consistent across these two aggregate statistics and two risk aversion assumptions, hereafter we use exponential utility as the default risk aversion assumption and equal-weighted average to aggregate the risk measures.

Figure 2.3(b) illustrates the sustainable risk pools generated. The visualization suggests that our optimal pooling results are geographically coherent — that is, producers within the same risk pool tend to be situated nearby. Additionally, it is noteworthy that some counties, despite being distant from one another, are grouped within the same risk pool. This phenomenon can be attributed to these counties possibly adhering to a shared data-generating process, as highlighted by Ker et al. (2015) and Tack and Ubilava (2015). To provide a more focused discussion and interpret the geographical implications of the risk pooling results, we perform a subsample analysis in Section 2.4.3.

### 2.4.3 Subsample Analysis: Geographical Implications

The objective of our focus on Illinois is twofold. First, Illinois is a major corn-producing state, contributing significantly to the U.S.’s total corn production. Second, the state exhibits a unique geographical profile. As shown in Figure 2.4(a), Illinois can be divided into three distinct geographical regions from north to south:

- Northern Illinois, including the Chicago Municipal Area and surrounding Charles

---

<sup>15</sup>The monetary calculations were based on the assumption that corns were sold on December 31, 2019, at the spot price. This spot price was derived from the corn commodity index as provided by FactSet Research Systems (CORNCI-FDS).

**TABLE 2.2**

**Baseline Results: Sustainable Risk Pooling Quality Analysis**

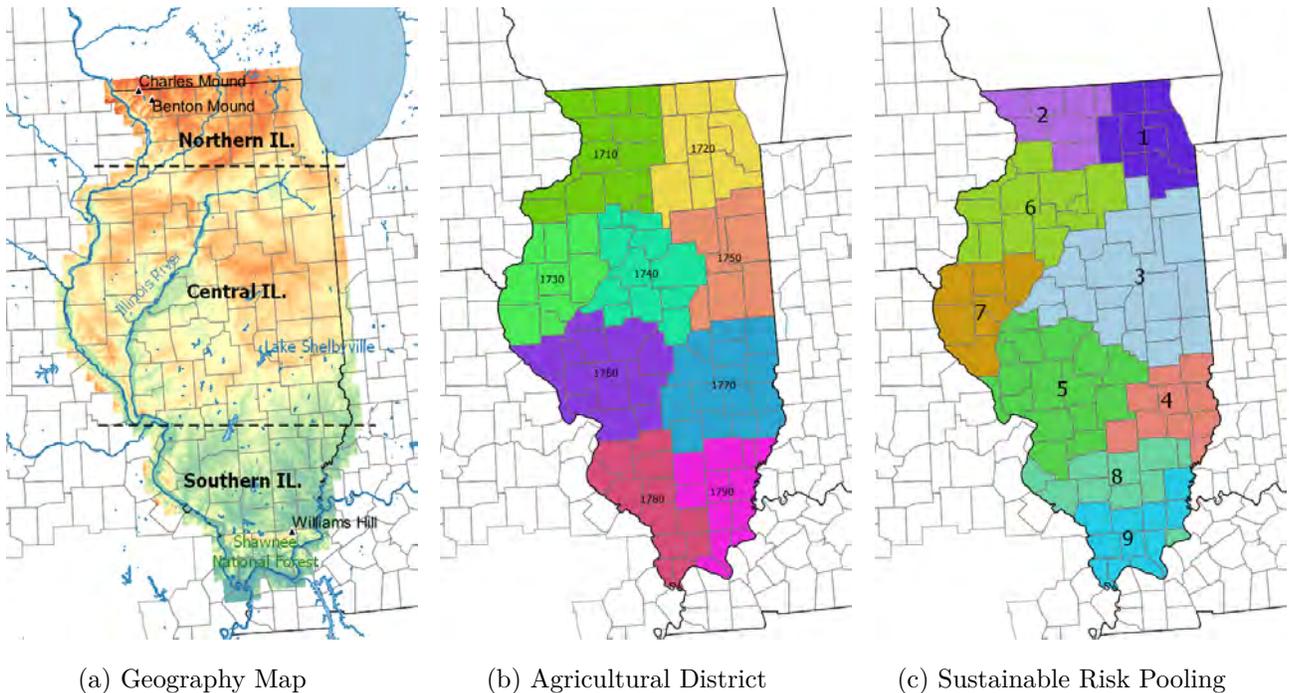
We evaluate the effectiveness of area-yield insurance derived from risk pools generated through our sustainable risk pooling method. These pools are constructed under two prevalent risk aversion assumptions: exponential utility ( $A = 0.5$ ) and power utility ( $\gamma = 2.0$ ). This table presents three panels summarising the improvements in area-yield insurance using sustainable risk pools compared to those based on agricultural districts and state boundaries. These improvements are quantified by basis risk (MSE) and tail risk (VaR and ES). We report results with risk measures aggregated through both equal weighting (Equal Wgt.) and cropland-size weighting (Land-size Wgt.) approaches. All results are based on a Sustainable Compliance Ratio (SCR) of 95%.

	Exponential Utility ( $K^* = 94$ )		Power Utility ( $K^* = 92$ )	
	Equal Wgt.	Land-size Wgt.	Equal Wgt.	Land-size Wgt.
<i>Panel A: Basis Risk Improvement - Mean Squared Error</i>				
MSE (1e-4)	5.31	5.15	5.26	4.86
Impv. vs. AGD	29.41%	21.65%	30.15%	26.08%
Impv. vs. State	58.46%	53.03%	58.90%	55.68%
<i>Panel B: Tail Risk Improvement - Value-at-Risk (5%)</i>				
VaR (1e-2)	5.17	5.83	5.13	5.51
Impv. vs. AGD	2.68%	8.02%	1.73%	2.01%
Impv. vs. State	6.04%	10.65%	5.05%	4.50%
<i>Panel C: Tail Risk Improvement - Expected Shortfall (5%)</i>				
ES (1e-2)	5.05	5.71	5.00	5.39
Impv. vs. AGD	3.26%	8.43%	2.28%	2.39%
Impv. vs. State	8.65%	12.89%	7.62%	6.61%
<i>Benchmarks</i>				
	AGD Boundaries		State Boundaries	
MSE (1e-4)	7.53	6.57	12.79	10.96
VaR (1e-2)	5.04	5.40	4.88	5.27
ES (1e-2)	4.89	5.26	4.64	5.05

Mountain.

- Central Illinois, known for its prairie landscapes and the Illinois River.
- Southern Illinois, distinguished by its warmer climate and location between the Mississippi and Ohio Rivers.

This geographical diversity offers valuable insights into our risk pooling results, providing a contextualized understanding within Illinois' diverse topography.



**Figure 2.4.** Illinois geography and risk pooling visualization.

To emphasize the geographical underpinnings of our findings, we illustrate the risk pooling outcomes for  $K = 9$  in Figure 2.4(c). Our analysis effectively partitions Illinois into its three major geographical regions: Northern Illinois, represented by Pools 1 and 2; Central Illinois, including Pools 3 to 7; and Southern Illinois, encompassing Pools 8 and 9. Within these broad categories, more specific distinctions are apparent. For example, Northern Illinois is divided into the Chicago municipal area (Pool 1) and the area around Charles Mountain (Pool 2). In Central Illinois, the Illinois River serves as a natural boundary between risk pools, with noticeable elevation differences observed between Pools 3, 4, and 5. Southern Illinois exhibits

a contrast between prairie landscapes in Pool 8 and the region with the higher forest coverage in Pool 9. These geographical insights are not captured in administrative-based risk pooling methods, such as those based on agricultural districts shown in Figure 2.4(b). Our risk pooling strategy not only achieves better basis risk and tail risk reductions but also aligns more closely with the geographical contours of the state.

#### 2.4.4 Performances of the Gap Statistic-based Framework

To further demonstrate the effectiveness of the proposed risk pooling framework, in this section, we compare our behavior-based approach with a widely accepted statistical method in identifying the optimal  $K^*$ , Gap statistic (Tibshirani et al., 2001). This statistical-based method involves the calculation of  $Gap(K)$ , the gap statistic, in choosing the optimal  $K^*$ . More specifically, consider  $Z$  as our primary dataset and  $Z^0$  as a dataset sampled from a uniform distribution, with both datasets having consistent ranges across each dimension. The gap statistic for  $K$  clusters,  $Gap(K)$ , is expressed as,

$$Gap(K) = E [\log(W(Z^0, K))] - \log(W(Z, K)). \quad (2.14)$$

Here  $W(Z, K)$  is the within-cluster dispersions for dataset  $Z$  and  $K$  clusters defined as

$$W(Z, K) = \sum_{k=1}^K \frac{1}{2N_k} \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} dist(i, j), \quad (2.15)$$

where  $dist(i, j)$  measures the Euclidean distance between  $i$  and  $j$ , and  $N_k$  is the cardinality of the  $k$ th risk pool. In practice,  $E [\log(W(Z^0, K))]$  can be estimated via Monte Carlo simulations.

Using the gap statistics, the value of  $K$  can be progressively increased until the difference between  $Gap(K + 1)$  and  $Gap(K)$  is not statistically significant. Tibshirani et al. (2001) introduce a 1-standard-error rule. Specifically, optimal  $K^*$  is identified as the first value of  $K$  for which the difference  $Gap(K + 1) - Gap(K)$  is less than or equal to the standard error of  $\log(W(Z^0, K))$ .

Table 2.3 displays area-yield insurance protection results from applying the gap statistic method to determine optimal  $K^*$ . The expected value,  $E[\log(W(Z^0, K))]$ , is estimated from 1000 simulations. Using this method, the optimal  $K^*$  is identified as 36, a value much lower than that suggested by our sustainable risk pooling approach. Notably, this  $K^*$  from the gap statistic method results in increased basis risk, highlighting its limitations in determining  $K^*$  for area-yield insurance risk pooling. More importantly, the  $K^*$  determined by gap statistic method lacks clear insurance or economic interpretations.

**TABLE 2.3**  
**Performances of the Gap Statistic-Based Risk Pooling**

We use the gap statistic to determine the optimal  $K^*$  for area-yield insurance. This table presents three panels summarising the improvements in area-yield insurance using sustainable risk pools compared to those based on agricultural districts and state boundaries. These improvements are quantified by basis risk (MSE) and tail risk (VaR and ES), with risk measures aggregated through an equal weighting approach.

Gap Statistics ( $K^* = 36$ )	
<i>Panel A: Basis Risk Improvement - Mean Squared Error</i>	
MSE (1e-4)	7.65
Impv. vs. AGD	-1.63%
Impv. vs. State	40.19%
<i>Panel B: Tail Risk Improvement - Value-at-Risk (5%)</i>	
VaR (1e-2)	5.03
Impv. vs. AGD	-0.16%
Impv. vs. State	3.10%
<i>Panel C: Tail Risk Improvement - Expected Shortfall (5%)</i>	
ES (1e-2)	4.87
Impv. vs. AGD	-0.37%
Impv. vs. State	4.83%

## 2.5 Robustness of Sustainable Risk Pooling

In this section, we examine the robustness of our proposed risk pooling approach by relaxing some assumptions made in the baseline results. Section 2.5.1 examines the impact of the sustainable compliance ratio (SCR) on our results. Subsequently, Section 2.5.2 investigates the effects of dimension reduction. Section 2.5.3 evaluates the sensitivity of assumptions on the cost function. In Section 2.5.4, we evaluate our optimal risk pooling method using three alternative basis risk measures from the area-yield insurance literature. Section 2.5.5

then allows producers to customize their protection ratios within various constraints. Finally, Section 2.5.6 discusses the integration of an equal cluster size constraint in the spectral clustering process, ensuring balanced risk pool sizes.

### 2.5.1 Impact of Sustainable Compliance Ratio

In the baseline results, we set the SCR at 0.95. This section explores the robustness of our results for varying levels of SCR. We assess SCR levels ranging from 0.85 to 1.0 and present the results in Table 2.4. We can see that as SCR rises, the optimal  $K^*$  declines from 135 to 43. Intuitively, reducing the number of  $K^*$  will increase basis risk, hence diminishing the effectiveness of the resulting area-yield insurance contract. Indeed, we see that relative to AGD, the improvement in producers' basis risk diminishes from 42.7% to 7.0% and the improvement in VaR and ES decreases from 4.3% to 0.4% and 5.3% to 0.3%, respectively. These findings confirm the negative association between the area-yield contracts effectiveness and the SCR, due to the trade-off between moral hazard and basis risk.

### 2.5.2 Impact of Retained Variance Ratio

While spectral clustering method effectively helps address the curse of dimensionality in the risk pooling process, the number of PCs retained in the algorithm has impact on the contract effectiveness. This section sensitivity tests the relationship between the number of PCs used in risk pooling and the associated basis risk, as shown in Figure 2.5. We can see that there is a trend emerging where the basis risk initially declines but eventually rises as the number of PCs increases. For our dataset, retaining a variance ratio between 85% and 95% from the original correlation matrix tends to produce risk pools with the lowest basis risk.

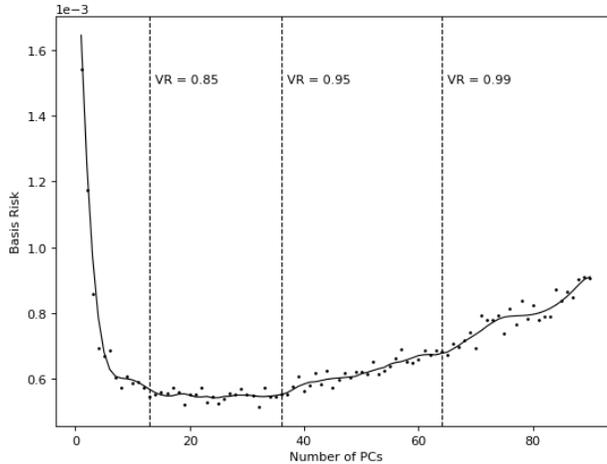
A significant implication from our analysis is that merely increasing the number of PCs for risk pooling does not guarantee optimal basis risk reduction, due to curse of dimensionality (Beyer et al., 1999). To illustrate, Table 2A.1 in Appendix 2A.2 investigates an extreme scenario where all the PCs from the spectral decomposition are used for risk pooling. The

**TABLE 2.4**

**Robustness Check: Impact of Sustainable Compliance Ratio Selection**

This table tests how the risk pooling results will be affected by the insurers' sustainable compliance ratio (SCR) selection. Producers are assumed to follow exponential utility with  $A = 0.5$ . The optimal  $K^*$  is presented in parentheses below each SCR selection. This table presents three panels summarising the improvements in area-yield insurance using sustainable risk pools compared to those based on agricultural districts (AGD) and state boundaries (State). These improvements are quantified by basis risk (MSE) and tail risk (VaR and ES), with risk measures aggregated through equal weighting.

	Sustainable Compliance Ratio Selection					
	1.0 ( $K^* = 43$ )	0.99 ( $K^* = 63$ )	0.98 ( $K^* = 73$ )	0.95 ( $K^* = 94$ )	0.9 ( $K^* = 115$ )	0.85 ( $K^* = 135$ )
<i>Panel A: Basis Risk Improvement - Mean Squared Error</i>						
MSE (1e-4)	7.00	6.26	5.77	5.31	4.83	4.32
Impv. vs. AGD	7.00%	16.79%	23.31%	29.41%	35.88%	42.65%
Impv. vs. State	45.27%	51.03%	54.87%	58.46%	62.27%	66.25%
<i>Panel B: Tail Risk Improvement - Value-at-Risk (5%)</i>						
VaR (1e-2)	5.06	5.09	5.14	5.17	5.22	5.26
Impv. vs. AGD	0.35%	1.07%	1.91%	2.68%	3.68%	4.33%
Impv. vs. State	3.64%	4.38%	5.25%	6.04%	7.07%	7.74%
<i>Panel C: Tail Risk Improvement - Expected Shortfall (5%)</i>						
ES (1e-2)	4.90	4.95	5.00	5.05	5.10	5.15
Impv. vs. AGD	0.29%	1.37%	2.37%	3.26%	4.42%	5.34%
Impv. vs. State	5.52%	6.66%	7.72%	8.65%	9.87%	10.84%



**Figure 2.5.** Basis risk of area-yield contracts using risk pools derived from various numbers of principal components.

analysis reveals a detrimental effect of curse of dimensionality on both basis risk and extreme risk improvements. In addition, no  $K^*$  value yields an ACR greater than 95% when all PCs are utilized in risk pooling.

### 2.5.3 Impact of Cost Ratio

In the baseline analysis, due to data availability constraints, we approximate producers' costs per acre as a fixed cost ratio  $C_{fixed}$  and a variable cost ratio  $C_{variable}$  relative to the expected revenue from production. To assess the sensitivity of our results to these important cost assumptions, we vary the fixed and variable costs by  $\pm 30\%$  and evaluate the performance of the sustainable risk pooling framework.<sup>16</sup>

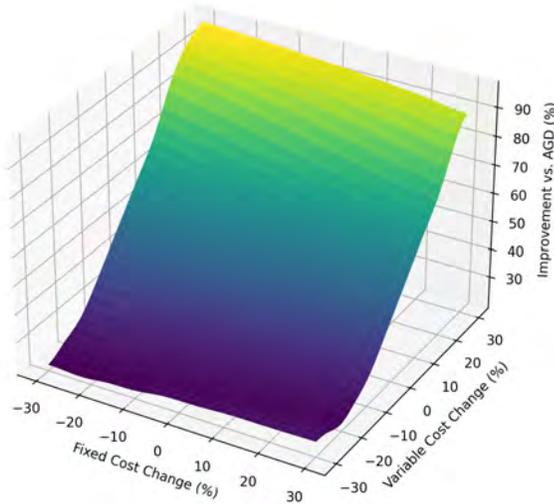
As shown in Figure 2.6, the proposed framework consistently mitigates basis risk across the tested range. In particular, we see less sensitivity in the results in fixed costs than variable costs. This is intuitive because variable costs have significant impact on producers' strategic decisions regarding the farming process. An increase in variable costs leads to lower expected marginal profits, prompting more producers, especially those with lower expected yields, higher variance, and less exposure to systemic risk (i.e., lower  $\beta_i$  in Equation 2.4), to adopt passive management strategies (Proposition 2.1 and 2.2). Consequently, as the variable costs rise, the size of risk pools will increase due to a heightening in passive management, which subsequently affects the basis risk improvement for the compliers.

### 2.5.4 Alternative Measures of Basis Risk

In our baseline analysis, we utilize Mean Squared Error (MSE) along with tail risk metrics, specifically Value-at-Risk (VaR) and Expected Shortfall (ES), to assess the effectiveness of risk pooling. This naturally leads to the question of how our methodology performs against other basis risk measures used in the area-yield literature. For example, (Miranda, 1991) use variance reduction to evaluate area-yield insurance programs, while (Deng et al., 2007) consider the correlation between index and individual yields. Additionally, (Elabed

---

<sup>16</sup>The increment is capped at 30% to ensure the marginal revenue remains non-negative.



**Figure 2.6.** Basis risk improvement in the sustainable risk pooling framework compared to AGD, measured against changes in fixed and variable costs.

et al., 2013) explore the likelihood of losses occurring without triggering payments. In this subsection, we apply these three alternative risk metrics to assess our sustainable risk pooling framework. Results presented in Table 2.5 confirms the robustness of our approach in improving basis risk across a range of measures.

### 2.5.5 Optimal Area-yield Contracts

In the baseline results, we assume that producers set *scale* of their policies at 100%. In this subsection, we allow producers to choose their optimal *scales*,  $\beta_i$ , in line with Equation (2.4) (Miranda, 1991; Mahul, 1999; Vercammen, 2000). We compare risk pooling results with different potential contractual constraints on these scales, as discussed by Barnett et al. (2005). Specifically, we examine three cases: In the first, producers select their optimal protection *scale* from a range of [90%, 110%]. The second case expands this range to [80%, 120%]. In the third case, producers are free to choose any value for their protection *scale*. Results are presented in Table 2.6. Comparing with our baseline results shown in Table 2.2, we can see

**TABLE 2.5**

**Robustness Check: Alternative Basis Risk Measures**

This table presents baseline results of the sustainable risk pooling method evaluated using three alternative basis risk measures. Panel A showcases variance reduction, as outlined by [Miranda \(1991\)](#). Panel B illustrates the correlation between individual and area yields, following the approach of [Deng et al. \(2007\)](#). Finally, Panel C delves into the probability of incurring losses without indemnity payments, based on the methodology proposed by [Elabed et al. \(2013\)](#).

Exponential Utility ( $K^* = 94$ )		
Equal Wgt.		
<i>Panel A: Yield variance reduction (<a href="#">Miranda, 1991</a>)</i>		
Var. Reduction	0.58	
Impv. vs. AGD	30.92%	
Impv. vs. State	106.58%	
<i>Panel B: Corr. inv. and area yield citecite (<a href="#">Deng et al., 2007</a>)</i>		
Corr. Inv. Y. & Area Y.	0.89	
Impv. vs. AGD	7.30%	
Impv. vs. State	26.35%	
<i>Panel C: Pr. of loss incurred wo payment (<a href="#">Elabed et al., 2013</a>)</i>		
Pr. Loss wo Pay	0.09	
Impv. vs. AGD	39.79%	
Impv. vs. State	60.92%	
<i>Benchmarks</i>		
	AGD Boundaries	State Boundaries
Var. Reduction	0.44	0.28
Corr. Inv. Y. & Area Y.	0.83	0.71
Pr. Loss wo Pay	0.16	0.24

that these variations in *scale* constraints have a minimal impact on risk improvements from our proposed sustainable risk pooling framework.

**TABLE 2.6**

**Robustness Check: Impact of Protection *scale***

This table summarizes the risk pooling results under scenarios where producers can choose their optimal *scale* within certain contract constraints. We examine three cases where *scale* ranges from [90%, 110%], [80%, 120%], and unconstrained *scale*.

	90% $\leq$ <i>scale</i> < 110% ( $K^* = 94$ )	80% $\leq$ <i>scale</i> < 120% ( $K^* = 94$ )	unconstrained <i>scale</i> ( $K^* = 94$ )			
<i>Panel A: Basis Risk Improvement - Mean Squared Error</i>						
MSE (1e-4)	5.11	5.06	5.10			
Impv. vs. AGD	28.48%	28.00%	26.77%			
Impv. vs. State	58.81%	58.63%	57.58%			
<i>Panel B: Tail Risk Improvement - Value-at-Risk (5%)</i>						
VaR (1e-2)	5.18	5.18	5.17			
Impv. vs. AGD	2.49%	2.50%	2.45%			
Impv. vs. State	6.04%	6.04%	6.01%			
<i>Panel C: Tail Risk Improvement - Expected Shortfall (5%)</i>						
ES (1e-2)	5.04	5.04	5.04			
Impv. vs. AGD	3.00%	3.00%	2.95%			
Impv. vs. State	8.62%	8.62%	8.58%			
<i>Benchmarks</i>						
	AGD	State	AGD	State	AGD	State
MSE (1e-4)	7.15	12.41	7.02	12.21	6.96	12.02
VaR (1e-2)	5.04	4.88	5.04	4.88	5.04	4.87
ES (1e-2)	4.89	4.64	4.89	4.64	4.89	4.64

## 2.5.6 Constructing Risk Pools with the Same Size

In our earlier discussions, we did not enforce any size restrictions on risk pools. Without such constraints, clustering algorithms can occasionally produce extremely small clusters. In the context of area-yield risk pooling, this might lead to the isolation of high-risk producers from the majority. As discussed in Section 2.3, diminutive risk pools can potentially encourage moral hazard. To mitigate this concern, in this subsection, we apply a uniform-size constraint using the algorithm proposed by Schubert et al. (2015), ensuring each cluster maintains a size of either ceiling or floor of  $N/K$ .

The results are summarized in Table 2.7. Comparing with our baseline results in Table 2.2, it is evident that the uniform size constraint has a larger value for the optimal  $K^*$ ,

while the enhancements in risk metrics remain largely consistent. In essence, with the same size constraint, our proposed method consistently outperforms the benchmark approaches.

**TABLE 2.7**  
**Robustness Check: Impact of a Same-Size Constraint**

This table shows the risk pooling results when sizes of all risk pools are restricted to be the same.

Exponential Utility ( $K^* = 90$ )	
Equal Wgt.	
<i>Panel A: Basis Risk Improvement - Mean Squared Error</i>	
MSE (1e-4)	5.33
Impv. vs. AGD	29.16%
Impv. vs. State	58.32%
<i>Panel B: Tail Risk Improvement - Value-at-Risk (5%)</i>	
VaR (1e-2)	5.16
Impv. vs. AGD	2.41%
Impv. vs. State	5.76%
<i>Panel C: Tail Risk Improvement - Expected Shortfall (5%)</i>	
ES (1e-2)	5.03
Impv. vs. AGD	2.98%
Impv. vs. State	8.36%

## 2.6 A Farm-level Analysis Based on Simulated Data

Due to the lack of access to farm-level production data, the main results of our empirical analysis are conducted using a county-level dataset, which represents the most detailed level of data available to us. To address the concerns that analyses at the county level might misinterpret individual risks, in this section, we apply the sustainable risk pooling framework to a simulated farm-level dataset, to further verify the performance of our proposed approach.

### 2.6.1 Farm-level Data Simulation

To proceed, we simulate the farm-level yield based on a risk factor-based model. More specifically, for producer  $i$  in year  $t$ , we assume her yield, denoted as  $y_{it}$ , can be attributed to four mutually independent factors:

$$y_{it} = \beta_{1i}f_{county,t} + \beta_{2i}f_{regional,t} + \beta_{3i}f_{macro,t} + \beta_{4i}f_{it}. \quad (2.16)$$

In this model,  $f_{county,t}$  represents the common risk factor specific to the county in which producer  $i$  operates. This factor captures unique county features that affect yield outcomes, such as local agricultural practices and technology supplies.  $f_{regional,t}$  represents inter-county common risk factors, including local characteristics influenced by geographical and climatic differences that affect individual yields.  $f_{macro,t}$  captures macro risk factors that apply to all producers, reflecting broader economic, policy, and environmental trends. Finally,  $f_{it}$  stands for producer  $i$ 's idiosyncratic risks, including individual production techniques and management practices.

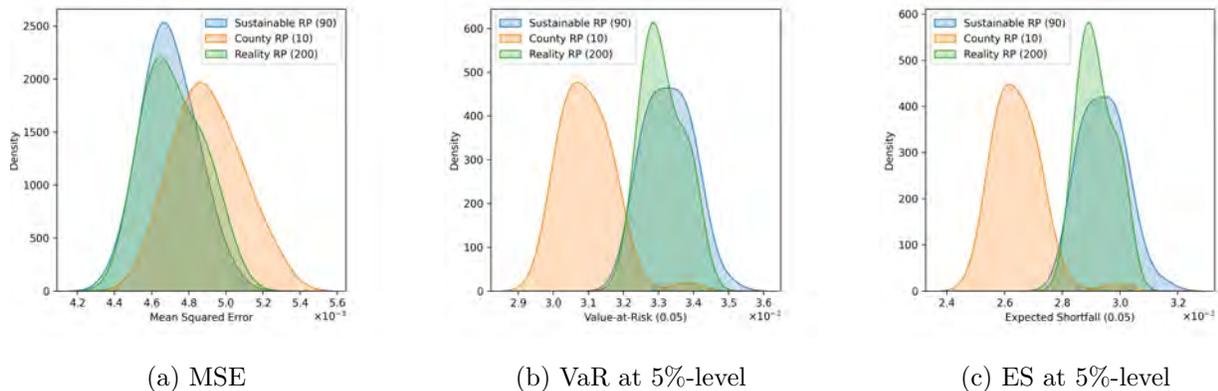
In the simulation procedure, the four risk factors for each producer are independently drawn from a standard normal distribution for each year. The parameters  $\beta_{1i}$  to  $\beta_{4i}$  represent the exposure of  $y_{it}$  to each of the four factors. In our simulation, we randomly draw each producer's  $\beta_{1i}$  to  $\beta_{4i}$  uniformly from  $[0, 1]$ , subject to the constraint that  $\sum_{j=1}^4 \beta_{ji} = 1$ . We also set the variance of each of the four factors to be equal to 1, meaning that the variance of  $y_{it}$  is also set to 1.

To improve the robustness of our simulation study, we conduct an ensembling procedure with a total of 50 simulation ensembles. In each ensemble, we simulate the production outcomes over 100 years for 10,000 producers. These producers are randomly assigned to one of 10 counties and 20 regional risk pools, resulting in a total of 200 distinct risk pools (denoted as (*Reality RP*)) in the simulation.

## 2.6.2 Performance of Sustainable Risk Pooling

We apply the sustainable risk pooling framework to each simulated ensemble, and on average, this method suggests an optimal  $K^* = 90$ . We use the county risk pools as a natural benchmark, which reflects the risk pooling for area-yield insurance in practice in the U.S.

Figure 2.7 displays the distribution of performance for both the sustainable risk pooling (*Sustainable RP*) and the county risk pooling (*County RP*) across simulated ensembles, particularly in terms of their effectiveness at capturing basis risk and tail risk as compared to actual risk pools. We can see from the figures that the sustainable risk pooling framework



**Figure 2.7.** Distribution of the performance metrics from different risk pooling methods on simulated farm-level data.

aligns more closely with reality risk pools than the county risk pool. This simulation study confirms the superiority of our proposed sustainable risk pooling framework, particularly in cases where there are inter-county common risk factors that cannot be adequately captured by arbitrary administrative boundaries.

## 2.7 Conclusion Remarks

In this chapter, we propose a behavior-based machine learning methodology for optimally determining area-yield insurance risk pools. This framework identifies the optimal number of risk pools  $K^*$  by analyzing the farming behaviors of producers with the protection of area-yield insurance contracts in a utility maximization framework. Then, the spectral clustering method is utilized to categorize producers into the  $K^*$  risk pools. This technique groups producers with similar production histories, enhancing the efficiency of the area-yield insurance contract and addressing high-dimensionality and computational complexity challenges.

Using county-level corn production data from the U.S. Heartland Region, we apply our proposed behavior-based machine learning risk pooling method. Empirical findings indicate that the proposed framework reduces contract basis risk and mitigates producers' tail risk. The approach proves robust across different models and contract parameters. When

compared to other statistical methods, our methodology provides improved risk reductions and yields meaningful geographical and economic insights, suggesting its potential as an alternative to the traditional area-yield insurance programs.

While the application in this chapter is mainly based on county-level data due to unavailability of farm-level observations, we extend the proposed methodology to simulated datasets of individual farms and reassess its efficiency. In addition, in this project, we assume a basic cost function to approximate producers' costs with two components of fixed or variable costs as proportions of their expected yields. This estimation can be further refined with more detailed data. It may also be valuable to evaluate the spatial and hedging efficiencies of various index-based contracts, such as area-yield insurance, weather derivatives, and satellite-driven vegetation index insurance, as referenced in (Woodard and Garcia, 2008; Zhu et al., 2018). We leave these topics for future research.

## 2A Appendix

### 2A.1 Proofs

To prove our results, we establish the following lemma.

**Lemma 2A.1.** *For producer  $i$  and  $j$ , both operating within the  $k$ th risk pool, the inequality  $EU(\mathbf{c}_i^A) - EU(\mathbf{c}_i^B) \geq EU(\mathbf{c}_j^A) - EU(\mathbf{c}_j^B)$  is a sufficient condition for  $\mathbf{1}(\mathbf{c}_i^*) \geq \mathbf{1}(\mathbf{c}_j^*)$ .*

Proof. The strategy choice of producer  $i$  is determined by the comparison of  $EU(\mathbf{c}_i^A)$  and  $EU(\mathbf{c}_i^B)$ , with an analogous criterion applied to producer  $j$ . Specifically, under the condition that  $EU(\mathbf{c}_i^A) - EU(\mathbf{c}_i^B) \geq EU(\mathbf{c}_j^A) - EU(\mathbf{c}_j^B)$ , one of the one of the following three scenarios will occur:

1. If both  $EU(\mathbf{c}_i^A) - EU(\mathbf{c}_i^B) \geq 0$  and  $EU(\mathbf{c}_j^A) - EU(\mathbf{c}_j^B) \geq 0$ , then both producers will choose active management strategies, leading to  $\mathbf{1}(\mathbf{c}_i^*) = \mathbf{1}(\mathbf{c}_j^*) = 1$ ;
2. If  $EU(\mathbf{c}_i^A) - EU(\mathbf{c}_i^B) \geq 0$  but  $EU(\mathbf{c}_j^A) - EU(\mathbf{c}_j^B) < 0$ , then Producer  $i$  will opt for optimal management while Producer  $j$  will adopt a passive approach. Consequently,  $\mathbf{1}(\mathbf{c}_i^*) = 1 > \mathbf{1}(\mathbf{c}_j^*) = 0$ ;
3. If both  $EU(\mathbf{c}_i^A) - EU(\mathbf{c}_i^B) < 0$  and  $EU(\mathbf{c}_j^A) - EU(\mathbf{c}_j^B) < 0$ , then both producers will choose passive management strategies, leading to  $\mathbf{1}(\mathbf{c}_i^*) = \mathbf{1}(\mathbf{c}_j^*) = 0$ .

Under the ceteris paribus condition, and without differentiating passive management strategies, we have  $EU(\mathbf{c}_i^B) = EU(\mathbf{c}_j^B)$ . To establish  $\mathbf{1}(\mathbf{c}_i^*) \geq \mathbf{1}(\mathbf{c}_j^*)$ , it is sufficient to show that  $EU(\mathbf{c}_i^A) \geq EU(\mathbf{c}_j^A)$ . Conversely, without differentiating active management strategies, proving  $\mathbf{1}(\mathbf{c}_i^*) \geq \mathbf{1}(\mathbf{c}_j^*)$  is sufficient through showing  $EU(\mathbf{c}_i^B) \leq EU(\mathbf{c}_j^B)$ .

**Lemma 2A.2.** *Let  $u$  be a utility function such that  $u'(x) > 0$  and  $u''(x) < 0$  for all  $x$ . Consider random wealth  $W_1$  and  $W_2$  with the same mean but  $\text{Var}(W_1) \leq \text{Var}(W_2)$ . It holds that  $E[u(W_1)] \geq E[u(W_2)]$ .*

Proof. Let  $w$  denote an individual's expected wealth, with  $W_1 = w + \epsilon_1$  and  $W_2 = w + \epsilon_2$ . Here  $\epsilon_1$  and  $\epsilon_2$  are two zero-mean random variables, with  $\mathbf{E}[\epsilon_1] = \mathbf{E}[\epsilon_2] = 0$  and  $\mathbf{Var}(\epsilon_1) \leq \mathbf{Var}(\epsilon_2)$ .

For notation simplicity, we denote the variance of  $\epsilon_1$  and  $\epsilon_2$  as  $\sigma_1^2$  and  $\sigma_2^2$ . Expanding  $u$  around  $w$  for both  $\mathbf{E}[u(W_1)]$  and  $\mathbf{E}[u(W_2)]$  as convergent Taylor series, we have

$$\begin{aligned} \mathbf{E}[u(W_i)] &= \mathbf{E}[u(w + \epsilon_i)] = \mathbf{E}[u(w) + \epsilon_i u'(w) + \frac{1}{2} \epsilon_i^2 u''(w) + O(\epsilon_i^3)] \\ &= u(w) + \frac{1}{2} u''(w) \sigma_i^2 + O(\mathbf{E}(\epsilon_i^3)), \quad i = 1, 2. \end{aligned}$$

Following [Pratt \(1964\)](#), we assume the third absolute central moment of  $\epsilon_1$  and  $\epsilon_2$  is of smaller order than the variance of  $\epsilon_1$  and  $\epsilon_2$ , respectively. Given that  $u''(\cdot) < 0$  and  $\sigma_1^2 \leq \sigma_2^2$ , under regularity conditions, we have

$$\mathbf{E}[u(W_1)] - \mathbf{E}[u(W_2)] = \frac{1}{2} u''(w) (\sigma_1^2 - \sigma_2^2) + o(\sigma_1^2 - \sigma_2^2) \leq 0.$$

The intuition of this lemma is that, given identical utility functions and levels expected wealth, a risk-averse individual dislike variance, i.e., the one with larger variance will experience a reduced expected utility.

**Proof of Proposition 1.** Ceteris paribus, the systemic component of the yield risk, represented by  $\mathbf{Var}(\beta_i^A (y_k - \mu_k))$ , increases as idiosyncratic component  $\sigma_{\epsilon_i^A}^2$  decreases. Thus, conclusions (a) and (b) are equivalent.

To prove that  $EU(\mathbf{c}_i^A) \geq EU(\mathbf{c}_j^A)$  under the condition of (a), we begin by substituting Equation (2.3) into the expected utility of active management. The expected utility of producer  $i$ 's strategy  $A$  is expressed as follows:

$$\begin{aligned} EU(\mathbf{c}_i^A) &= \mathbf{E} \left[ u \left( w_0 + y_i^A - C_i^A + I_k \right) \right] \\ &= \mathbf{E} \left[ u \left( w_0 + \mu_i^A + \beta_i^A (y_k - \mu_k) + \epsilon_i^A - (C_i^A - \mathbf{E}[I_k]) + (I_k - \mathbf{E}[I_k]) \right) \right] \quad (2A.1) \\ &= \mathbf{E} \left[ u \left( w_0 + \mu_i^A - (C_i^A - \mathbf{E}[I_k]) + Z_i \right) \right]. \end{aligned}$$

In this equation, a zero-mean random variable  $Z_i$  is defined as  $Z_i = \beta_i^A (y_k - \mu_k) + \epsilon_i^A + (I_k - \mathbf{E}[I_k])$ .

$E[I_k]$ ), and the net premium  $E[I_k]$  is separated from the total costs  $C_i^A$ . The variance of  $Z_i$  is given by

$$\begin{aligned}\text{Var}(Z_i) &= \text{Var}[\beta_i^A (y_k - \mu_k) + \epsilon_i^A] + \text{Var}[I_k - E[I_k]] + \text{Cov}[\beta_i^A (y_k - \mu_k) + \epsilon_i^A, I_k - E[I_k]] \\ &= \sigma_{y_i^A}^2 + \text{Var}(I_k) + \beta_i^A \text{Cov}(y_k, I_k).\end{aligned}$$

Recall that  $\text{Cov}(y_k, I_k) < 0$ , since the area-yield indemnity  $I_k$  is effectively a put option on area-yield  $y_k$ . With  $\sigma_{y_i^A}^2 = \sigma_{y_j^A}^2$  and  $\text{Cov}(y_k, I_k) < 0$ , the condition  $\beta_i^A \geq \beta_j^A$  leads to  $\text{Var}(Z_i) \leq \text{Var}(Z_j)$ . Ceteris paribus, Lemma 2A.2 imply that  $EU(\mathbf{c}_i^A) \geq EU(\mathbf{c}_j^A)$ , and combining Lemma 2A.1, we have  $\mathbf{1}(\mathbf{c}_i^*) \geq \mathbf{1}(\mathbf{c}_j^*)$ .

**Proof of Proposition 2.** From the expected utility of active management (Equation (2A.1)), we obtain the following comparative static derivatives

$$\frac{\partial E[u(\mathbf{c}_i^A)]}{\partial \mu_i^A} = E[u'(\mathbf{c}_i^A)] > 0, \quad \frac{\partial E[u(\mathbf{c}_i^A)]}{\partial C_i^A} = -E[u'(\mathbf{c}_i^A)] < 0,$$

where  $u'(\mathbf{c}_i^A)$  represents the marginal utility of active management. Therefore,  $E[u(\mathbf{c}_i^A)] \geq E[u(\mathbf{c}_j^A)]$  when  $\mu_i^A \geq \mu_j^A$  or  $C_i^A \leq C_j^A$ , validating (a) and (c). (b) and (d) can be verified in a similar fashion by checking the marginal utilities of passive management.

## 2A.2 Additional analyses

### Curse of dimensionality

In this section, we test the impact of curse of dimensionality. In particular, we want to see how the risk pooling result will be affected if all the PCs are used in risk pooling. We can see from Table 2A.1 that the optimal risk pooling framework is not effective at all due to curse of dimensionality. In addition, no  $K$  value yields an ACR greater than 95% when all PCs are utilized in risk pooling.

**TABLE 2A.1**  
**Impact of High Dimensionality**

This table tests how the risk pooling result will be affected if all the PCs are used in risk pooling. Producers are assumed to follow exponential utility with  $A = 0.5$ . Each column shows the selection of a value of  $K$  and the attained compliance ratio (ACR) is presented below. This table presents three panels summarising the improvements in area-yield insurance using sustainable risk pools compared to those based on agricultural districts and state boundaries. These improvements are quantified by basis risk (MSE) and tail risk (VaR and ES), with risk measures aggregated through equal weighting.

	Selection of the Number of Risk Pools ( $K$ )		
	$K = 30$ (ACR = 94.57%)	$K = 60$ (ACR = 88.56%)	$K = 90$ (ACR = 82.94%)
<i>Panel A: Basis Risk Improvement - Mean Squared Error</i>			
MSE (1e-4)	17.59	17.54	17.65
Impv. vs. AGD	-133.68%	-133.07%	-134.51%
Impv. vs. State	0.08%	0.08%	0.08%
<i>Panel B: Tail Risk Improvement - Value-at-Risk (5%)</i>			
VaR (1e-2)	4.70	4.70	4.70
Impv. vs. AGD	-6.74%	-6.72%	-6.74%
Impv. vs. State	-3.69%	-3.67%	-3.69%
<i>Panel C: Tail Risk Improvement - Expected Shortfall (5%)</i>			
ES (1e-2)	4.38	4.37	4.38
Impv. vs. AGD	-10.34%	-10.48%	-10.36%
Impv. vs. State	-5.66%	-5.80%	-5.68%

## Parametric approach to determine producers' utility

While we have previously explored the performance of the proposed area-yield insurance using a nonparametric framework, in this analysis, we delve into an alternative approach. Here, the insurer employs parametric models to estimate the expected utility of producers.

We employ a multivariate normal distribution to illustrate the parametric approach. In a manner analogous to the baseline scenario, we consider a representative producer, denoted as  $i$ , who faces a choice between an active production management strategy and a passive one.

For the scenario in which she adheres to active management, we fit her historical individual yields,  $y_i$ , and the area-yields,  $y_k$ , to a bivariate normal distribution. From this distribution, we simulate  $N$  random samples of  $(y_i, y_k)$ . Using 1,000 simulation iterations, her expected utility is subsequently estimated through Equation (2.8).

In contrast, if she opts against active management, her historical individual yield is set

to zero, and the area-yield is adjusted to  $y_k^0 = y_k(1 - \pi_i)$ . Here,  $\pi_i$  represents the proportion of her cropland size relative to the total cropland size of her associated risk pool. Following this, a univariate normal distribution is fitted to  $y_k^0$ , and we employ the derived model to simulate  $(y_i^0, y_k^0)$ . The subsequent procedures mirror those of the active management case. The insurer then determines her behavior by comparing her expected utility across these two scenarios, as shown in Equation (2.9).

The risk pooling results using this parametric approach are summarized in Table 2A.2. Comparing with Table 2.2, we can see the optimal  $K^*$  and the improvements in risk metrics are about the same level as our baseline results.

**TABLE 2A.2**  
**Sustainable Risk Pooling Results with Parametric Utilities**

This table shows risk pooling results using parametric models. Producers are assumed to follow exponential utility with  $A = 0.5$ . This table presents three panels summarising the improvements in area-yield insurance using sustainable risk pools compared to those based on agricultural districts and state boundaries. These improvements are quantified by basis risk (MSE) and tail risk (VaR and ES), with risk measures aggregated through equal weighting.

Exponential Utility ( $K^* = 90$ )	
Equal Wgt.	
<i>Panel A: Basis Risk Improvement - Mean Squared Error</i>	
MSE (1e-4)	5.39
Impv. vs. AGD	28.39%
Impv. vs. State	57.86%
<i>Panel B: Tail Risk Improvement - Value-at-Risk (5%)</i>	
VaR (1e-2)	5.16
Impv. vs. AGD	2.40%
Impv. vs. State	5.75%
<i>Panel C: Tail Risk Improvement - Expected Shortfall (5%)</i>	
ES (1e-2)	5.03
Impv. vs. AGD	2.94%
Impv. vs. State	8.31%

# Chapter 3

## From Meteorology to Market: A Geo-Hierarchical Deep Learning Approach for Flood Risk Pricing

### 3.1 Introduction

Floods are the most frequent natural disasters, accounting for one-third of fatalities from natural calamities since 2011 (Swiss Re, 2022). In the context of climate change, the frequency and severity of flood events have escalated significantly, driven by rising sea levels, intensified precipitation patterns, and erosion from the degradation of natural flood protection systems like water meadows and coastal mangroves. This escalation poses substantial pressure on global risk reduction and risk transfer mechanisms (World Economic Forum, 2023). The economic and human impacts are equally severe, with billions of dollars in damages and significant loss of lives annually (Boudreault et al., 2019). For instance, the World Economic Forum has highlighted floods as one of the top global risks, affecting 1.8 billion people (i.e., 23% of the world population), necessitating advanced methodologies for accurate risk assessment and mitigation. In 2022 alone, flood risk cost the global economy more than \$82 billion (Swiss Re, 2022).

Well-devised risk management, including insurance, is essential for climate adaptation,

protecting individuals, businesses, and communities from the financial impacts of flooding. Despite its importance, the flood insurance market suffers from low coverage. Rising flood losses already compel insurance companies to increase their capital base, potentially leading to prolonged periods of unprofitability. Uninsured risks remain a significant concern, as inadequate financial resources for relief and recovery adversely affect society, the economy, and the well-being of people (Jongman et al., 2014). The insurance gap has grown significantly, from \$117 billion in 2020 to \$161 billion in 2021. While nearly 29% of the global population is exposed to flood risks, only 7% of flood losses in emerging markets and 31% in developed economies were insured over the two decades (Swiss Re, 2022). In the U.S., despite the National Flood Insurance Program (NFIP), the coverage of flood damages remains fairly low (Kousky et al., 2017). The U.S. Congress mandates flood insurance for properties in designated 100-year floodplain with federally-backed mortgages, yet recent flooding in Florida revealed that many at-risk properties remain uninsured (Flavelle, 2022; Santana and Phillis, 2022; Tolan and Devine, 2022). This gap highlights the need for enhanced awareness and accessibility of flood insurance. What makes things worse is that while risk management generally reduces the impacts of floods, it struggles with unprecedented events, which are becoming more frequent owing to climate change (Kreibich et al., 2022).

Accurate models serve as an essential foundation for policymakers to design resilient infrastructures and implement proactive mitigation strategies for flood risks. Accurate flood risk prediction enhances the resilience of communities and economies, reducing the overall socioeconomic impact of flooding events (Surminski et al., 2015). For the insurance industry, precise flood risk assessments underpin the development of appropriate insurance products, inform pricing strategies, and ensure the financial stability of insurance providers (Kousky and Kunreuther, 2018). Moreover, a robust framework for flood risk modeling that is both calibratable and transferable across different markets is also critical for flood insurance development.

Despite their importance, developing accurate flood risk models presents significant challenges. Traditional statistical models often fail to capture the complex interactions between hydrological and meteorological factors that influence flood events. The limited spatial

and temporal resolution of data also results in substantial prediction inaccuracies (Bouwer, 2013). Moreover, the dynamic nature of climate change introduces additional variability, further complicating reliable flood risk prediction. Physical models for catastrophe modeling, which simulate the mechanisms of hazards, are costly and slow to build and maintain. As a result, they are often largely unavailable for the majority of flood-prone areas, even in developed markets (Swiss Re, 2021). For example, the U.S. National Precipitation Expectations Standard, known as Atlas 14, developed by the National Oceanic and Atmospheric Administration (NOAA), found that over half of the U.S. population now lives in areas that are twice as likely to experience a severe “1-in-100-year flood” event as expected from Atlas 14 (Eby, 2023). While NOAA plans to update these standards with Atlas 15, the revisions are not expected until 2027 (NOAA, 2023). Inaccuracies in flood risk modeling not only undermine program effectiveness but also lead to limited public awareness, affecting the willingness to pay for insurance (Thistlethwaite et al., 2020).

The objective of this chapter is to address these challenges by developing an innovative Geo-Hierarchical Deep Learning (GHDL) framework specifically designed for flood risk prediction and flood insurance pricing. This framework leverages high-resolution meteorological and hydraulic data within a uniquely coded geographical structure. The model integrates the geographical connectivity of critical locations such as cities, dams, levees, reservoirs, harbors, and river crossings into the deep learning structure, thereby enhancing its relevance and transferability in varied physical contexts and anthropogenic conditions.

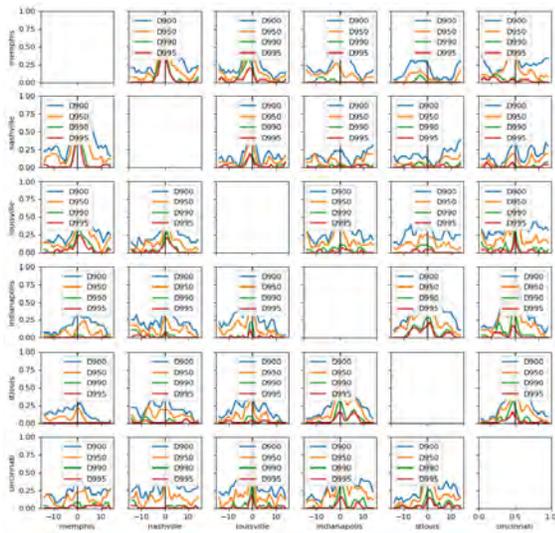
In the GHDL framework, we employ a Convolutional Neural Network (CNN) system for pluvial risk and a Wavelet Neural Network (WNN) system for fluvial risk, each integrating high-resolution climate and hydraulic data. The CNN layers can extract intricate information by processing image data, allowing the model to use more detailed meteorological data for enhanced training efficiency (Li et al., 2022). Conversely, the WNN layers excel in handling multiresolution and non-stationary time series information, making them effective for flood risk assessment techniques in the literature (Mosavi et al., 2018; Shafaei and Kisi, 2016). This dual system allows our model to understand the relationship between inland flood risks and the climatic conditions that contribute to both fluvial and pluvial floods concurrently. To

incorporate geographical information, at each designated site, the GHDL structure processes local data and then combines it with information from upstream locations. This integration provides predictive factors for downstream areas, thereby enhancing the accuracy of flood risk modeling.

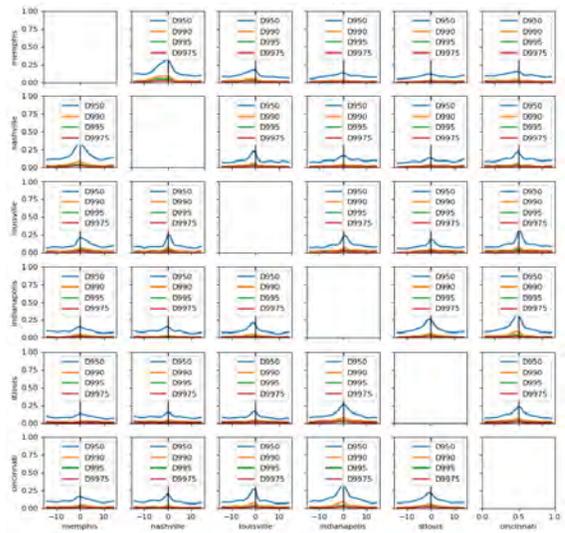
The GHDL structure has three key advantages. First, it enhances the model’s interpretability by integrating both local and upstream data, providing more transparent insights into the basis of its predictions. Second, this model is highly effective in handling spatio-temporal data, which is crucial for improving climate-related forecasts and developing effective pricing strategies. Flood risk exhibits strong spatial and temporal correlations, especially in tail-risk scenarios, as shown in Figure 3.1(a). We can see that this correlation becomes less pronounced in moderate-risk situations (Figure 3.1(b)). These findings suggest that the utilization of spatio-temporal data can improve the flood risk prediction accuracy by enhancing the model predictability in extreme scenarios. Third, the model effectively mitigates the risk of overfitting, a common issue in complex deep learning models with a large number of parameters. By introducing a geographical structural constraint, the GHDL model achieves the best bias-variance tradeoff, ensuring that the inclusion of spatial data not only enhances the model’s precision but also maintains its generalizability.

Utilizing the Mississippi River data, we provide compelling evidence that our proposed GHDL model outperforms its natural machine learning model benchmarks in forecasting flood risk and determining accurate pricing. Our empirical analysis reveals that in cities with upstream regions, the improvement in flood risk forecasting accuracy, as measured by the out-of-sample relative root mean squared error (RRMSE), can be as high as 12.5% to 16.5% in the most severe scenarios. The capability of the proposed GHDL model to provide more accurate predictions for extreme flood events is particularly valuable, given the documented inefficiency in traditional risk assessment models in extreme climate scenarios (Kreibich et al., 2022).

We further apply the calibrated model in pricing NFIP insurance policies. The results demonstrate substantial enhancements in net premiums and solvency capital requirements. In particular, the GHDL model consistently provides the most accurate predictions for claims



(a) Extreme flood risk scenarios



(b) Moderate flood risk scenarios

**Figure 3.1.** Spatio-temporal flood risk correlation across a cluster of six cities in the Mississippi River basin

Panel (a) plots the extreme flood risk scenarios (the most extreme 0.25% cases) and Panel (B) plots the moderate flood risk scenarios (the most extreme 10% cases). For example, the plot  $(x, y)$  in panel (a) shows if there is a flood event that exceeds the 0.25% threshold in city  $x$  at day 0, what is the probability that city  $y$  has a flood event happen between day  $[-14, 14]$ .

in both in-sample and out-of-sample data. Moreover, the pricing analysis shows that incorporating spatial and hierarchical structures significantly enhances flood risk pricing, with weather information playing a critical role. In addition, the GHDL model reduces the solvency capital requirement by 32.3%, when compared to benchmark ratemaking methods that overlook spatial and temporal flood risk factors.

The scope of this research is focused on the Mississippi River for two reasons. First, the availability of the NFIP datasets allows us to test the performance of the proposed GHDL model on a policy level. Second, the city clusters in the middle of the Mississippi River basin, with clear geographical connectivity, provide an excellent laboratory to demonstrate the GHDL model.

The time scope of our research covers several major flood events that occurred along the Mississippi River between 2009 and 2020. Significant floods during this period include the 2009 flood, which primarily impacted North Dakota, South Dakota, Minnesota, and Iowa, causing widespread damage to properties and infrastructure. The 2011 flood, one of the most significant in recent history, affected Missouri, Tennessee, Arkansas, Mississippi, and Louisiana, resulting in extensive agricultural and infrastructural damage. The 2013 flood, triggered by heavy rains and snowmelt, significantly impacted Illinois, Missouri, and Iowa. The severe flood from late December 2015 to early January 2016 primarily affected Missouri and Illinois, displacing residents and causing considerable damage. Lastly, the 2019 flood, one of the longest-lasting, affected Iowa, Illinois, Missouri, Arkansas, Mississippi, and Louisiana, leading to extensive agricultural losses and property damage. These events underscore the frequent and severe nature of flooding in the Mississippi River basin, highlighting its critical importance for flood risk research and management.

The discussion of flood risk modeling and its social impact has been extensively covered in engineering and social science literature over the past two decades (Lechowska, 2018). A diverse array of machine learning techniques has been scrutinized, including artificial neural networks (Li et al., 2009; Wu et al., 2009; Abbot and Marohasy, 2014), classification and regression trees (Dehghani et al., 2017; Choubin et al., 2018), support vector machines (Sang et al., 2008), and wavelet neural networks (Kumar et al., 2015; Seo et al., 2015; Shafaei and

Kisi, 2016). More recently, advanced deep learning techniques such as convolutional neural networks and recurrent neural networks (Guo et al., 2020; Zhang and Xiao, 2000) have been explored.<sup>1</sup> However, the engineering literature tends to focus on short-term horizons, typically limited to a single day, which is suitable for risk reduction and relief but not for insurance pricing. Moreover, it lacks methodologies for estimating the financial impact of flood risks.

Another stream of literature studies flood risk mitigation from an actuarial and insurance perspective. For instance, Boudreault et al. (2019) explore private insurance pricing and spatial segmentation of flood risk using a hierarchical model, enabling individual-level risk assessment. In contrast, Browne et al. (2018) identify a negative correlation between county-wide housing development in Florida and participation in the mandatory NFIP. Additionally, Wing et al. (2020) conduct a comprehensive analysis of the NFIP claim dataset, challenging the conventional assumption that flood damage increases monotonically with rising water depth. Their findings suggest that flood losses more accurately follow a beta function. Collier et al. (2022) analyze households' flood insurance coverage decisions recorded in the NFIP, finding that consumers often fully insure their homes despite high premium loads, influenced by industry practices and probability distortions in decision-making. More recently, Tesselaar et al. (2022) and Hossain et al. (2022) examine the demand and coverage gaps in flood insurance between developed regions, like Europe, and developing areas, such as Bangladesh, in the context of evolving climate conditions. Furthermore, Boudreault and Ojeda (2022) emphasize that the challenges in achieving risk differentiation at the homeowner level can be mitigated by forming homogeneous ratemaking groups. They introduce a clustering technique to establish these risk pools and thoroughly examine the potential problem of adverse selection. Conell-Price et al. (2022) demonstrate that behavioral interventions, such as default enrollment in supplemental flood insurance, significantly increase take-up rates, highlighting the potential of auto-enrollment to enhance financial resilience against flood risks. From a financial perspective, Hu (2022) highlights the influence of peer effects on flood insurance decisions, estimating that an individual's likelihood of purchasing flood insurance increases by 1-5% when a distant connection experiences a flood event. Lastly, Boonen et al.

---

<sup>1</sup>Further details on these advancements can be found in Mosavi et al. (2018).

(2024) study the risk-sharing mechanism of the centralized insurance market and apply the theory to the NFIP, finding clear benefits in risk aggregation and geographical diversification.

The GHDL model presented in this chapter makes an important contribution to the literature by integrating geographic hierarchical structures with advanced deep learning techniques. This approach not only enhances the accuracy of flood risk predictions but also facilitates the development of dynamic insurance products. By providing a robust predictive model, the GHDL framework offers significant theoretical and practical implications, particularly in improving risk assessment and insurance pricing strategies in the context of a changing climate.

The rest of the chapter proceeds as follows. Section 3.2 introduces the GHDL model proposed in this chapter. Section 3.3 illustrates the prediction performance of the GHDL model in empirical flood risk modeling. Section 3.4 applies the prediction model to price NFIP policies. Section 3.5 addresses the interpretability of deep learning models. Section 3.6 concludes.

## 3.2 A Geo-Hierarchical Deep Learning Model

Flood risk forecast models are the foundation of flood insurance pricing. In the U.S., daily river flow has emerged as an important and promising indicator for flood risk, and accurate predictions can provide critical information for early warning systems for flood risk (Quinn et al., 2019). Therefore, we utilize daily river flow measurements in our analysis. In particular, we propose a Geo-Hierarchical Deep Learning (GHDL) framework designed for flood insurance pricing through predicting river flow. This structure combines hydraulic and meteorological data, organizing it hierarchically based on geographical locations to enhance predictive accuracy.

Let  $Q_{it}$  represent the daily river flow at gauging site  $i = 1, \dots, I$  at time  $t = 1, \dots, T$ .  $Q_{it}$  quantifies the volume of water passing through location  $i$  in a river per unit of time.<sup>2</sup> It is well

---

<sup>2</sup> $Q_{it}$  is also known as water discharge, and is measured in cubic feet per second ( $ft^3/s$ ) according to the United States Geological Survey (USGS) historical land water dataset.

documented in the literature that water flow tends to be non-stationary and exhibits spatial and temporal dependencies, often correlating with hydraulic and meteorological information (Boudreault et al., 2019). Our goal is to forecast the river flow in period  $t$ ,  $Q_{it}$ , based on information from previous time periods up to  $t - 1$ . More specifically,

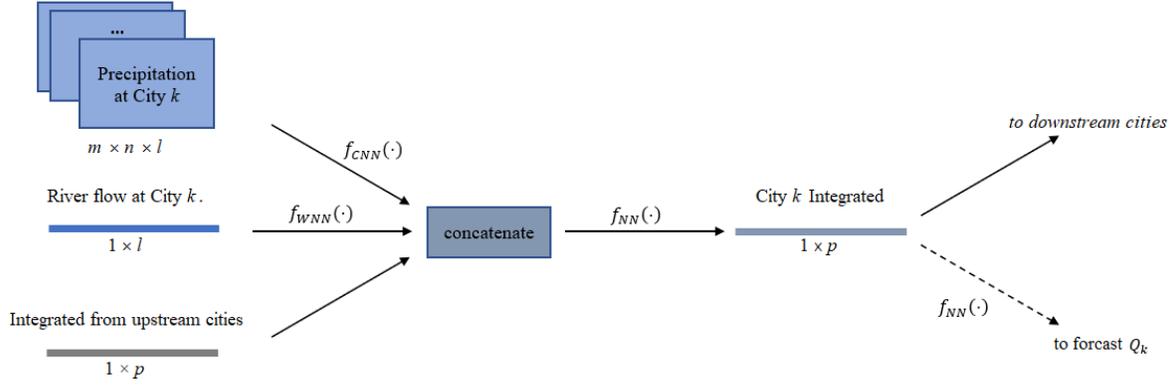
$$\hat{Q}_{it} = \mathbf{E}[Q_{it} | \mathbf{Q}_{t-1}, \mathbf{W}_{t-1}] = f(\mathbf{Q}_{t-1}, \mathbf{W}_{t-1}), \quad (3.1)$$

where  $\mathbf{Q}_{t-1} = \{Q_{it} : i = 1, \dots, I, t = t - 1, \dots, t - l\}$  and  $\mathbf{W}_{t-1} = \{W_{it} : i = 1, \dots, I, t = t - 1, \dots, t - l\}$  comprise the river flow and weather information sets for all gauges from  $t - l$  to  $t - 1$ , respectively. We denote the mean estimation of the constructed deep learning model as  $f(\cdot)$ .

### 3.2.1 Model Architecture

Inland flood risks within a specified region are generally categorized into two types: fluvial and pluvial floods. Fluvial floods are primarily influenced by river flow dynamics upstream, while pluvial floods result from intense rainfall. The simultaneous occurrence of both types can significantly amplify the impacts of each (Chen et al., 2010). Our study aims to explore the relationship between inland flood risks and the climatic conditions that contribute to both fluvial and pluvial floods concurrently.

To achieve this, we propose a Geo-Hierarchical Deep Learning (GHDL) structure that employs a dual-system methodology. Figure 3.2 displays the information integration process of the proposed GHDL model. Precipitation data, relevant to pluvial flood analysis, are processed using a Convolutional Neural Network (CNN) framework. Conversely, river flow data, crucial for fluvial flood analysis, are processed using a Wavelet Neural Network (WNN) system. This bifurcated approach allows for a detailed understanding of the distinct mechanisms driving each type of flood. This locally processed information is subsequently merged with data integrated from upstream cities, creating a comprehensive information set for local flood risk forecasting. This integrated dataset is not only utilized for local assessments but is also forwarded to downstream cities.

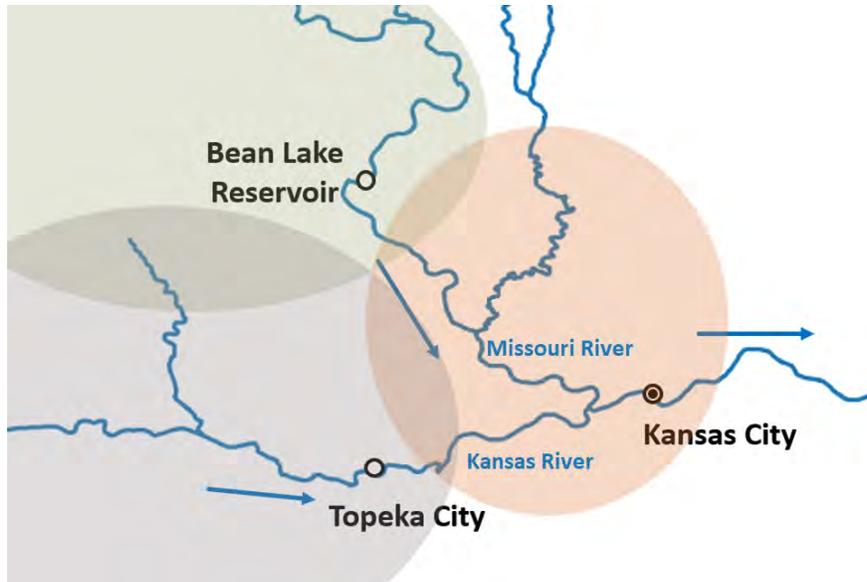


**Figure 3.2.** Information integration process of GHDL structure

This figure illustrates the information integration process of GHDL structure. For a given city  $k$ , the meteorological data (e.g., precipitation) are processed by a CNN system, while the hydraulic data (river flow) are processed by a WNN system. The processed meteorological and hydraulic data are then combined with the integrated information obtained from the upstream cities. The resulting data are fed into a NN system to generate the integrated output for City  $k$ .

The geographical hierarchy of cities (or locations) within a river basin is determined by their connectivity along the river channel. As an illustrative example, Figure 3.3 demonstrates the geographical hierarchy among Kansas City, Topeka City, and Bean Lake Reservoir. Kansas City is situated at the confluence of the Kansas River and the Missouri River, indicating that its flood risk is influenced by the meteorological and hydraulic conditions upstream of both rivers.

To forecast flood risk in this cluster, our GHDL structure encodes the geographical connectivity of the three cities or locations into the model’s architecture, as illustrated in Figure 3.4. More specifically, the integrated data from Bean Lake Reservoir and Topeka City are fed forward to combine with local information from Kansas City. Concurrently, the observed ground truth at Kansas City is utilized in a feedback mechanism to refine the training of flood prediction submodels in upstream cities or locations, thereby forming an integrated information set for Kansas City. The model is trained through global optimization, using observed flood risks at all cities or locations as ground truth. This model offers three immediate advantages. First, the incorporation of connectivity into the model enhances its interpretability. By utilizing upstream information for predictions, we gain a clearer understanding of how the model formulates its forecasts. Second, the encoded connectivity



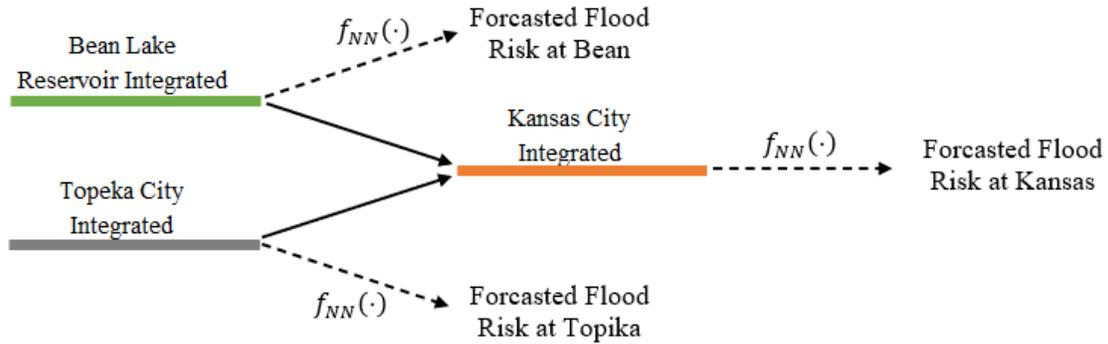
**Figure 3.3.** Geographic connectivity around Kansas city

This figure illustrates the hierarchical geographical locations between Kansas City, Topeka City (located on Kansas River), and Bean Lake Reservoir (Located on Mississippi River).

streamlines the optimization process. It narrows the search space for the optimization algorithm, leading to faster convergence times and making the model more efficient. Third, the integration of connectivity acts as a constraint, simplifying the model, which otherwise might be overly complex due to a large set of parameters. Hence, this approach effectively reduces the risk of overfitting in such an intricate model.

### 3.2.2 Convolutional Neuron Network Model

CNNs are a specialized type of deep learning neural network widely utilized in image recognition and computer vision tasks. We use CNNs to handle the pluvial segment of our deep learning framework, as they can process image precipitation data. Compared to aggregating precipitation at a certain level, image precipitation data provides high spatial resolution, capturing detailed variations in precipitation patterns across a geographical area. Due to local landscape variation, heavy rainfall at some locations can cause higher flood risks than others. Such patterns may be obscured in aggregated precipitation data but can be captured by image precipitation data.



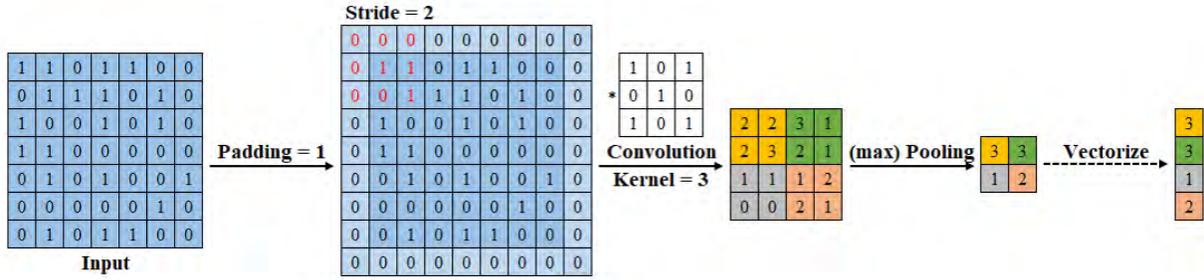
**Figure 3.4.** The GHDL structure for Kansas city

This figure demonstrate how the the geographic connectivities between Kansas City, Topeka City and bean Lake Reservoir will be coded into the GHDL structure.

The architecture and capabilities of CNNs have been profoundly influenced by several key developments in the field. The foundational paper by [LeCun et al.](#), introducing the LeNet-5 model, set the stage for the use of CNNs in document recognition. Building on this groundwork, [Krizhevsky et al.](#) in 2012 introduced AlexNet, a deeper CNN architecture that significantly advanced image classification on the ImageNet dataset. This breakthrough highlighted the substantial capabilities of deep learning in computer vision. Subsequent enhancements in CNN design have further refined the model (see, e.g., [Simonyan and Zisserman, 2014](#); [Szegedy et al., 2015](#); [Ronneberger et al., 2015](#); [He et al., 2016](#); [Redmon et al., 2016](#), among others).

CNNs offer several advantages: (1) Local connections: Instead of each neuron in a layer being connected to all neurons in the previous layer, it is connected to only a small number of neurons. This reduces the number of parameters and speeds up convergence. (2) Weight sharing: A group of connections can share the same weights, reducing the number of parameters even further. (3) Down-sampling: The pooling layer uses the principle of local correlation to down-sample an image, reducing the amount of data while retaining useful information ([Li et al., 2022](#)).

A typical CNN system comprises four main components. The “convolution” process is



**Figure 3.5.** Illustration of padding, stride, convolution, and pooling procedure

This figure demonstrate the procedure of a 2-dimensional CNN system. Here, we take the padding = 1, stride = 2, convolutional kernel size = 3, and max pooling = 2. In the last step, the convolution output is vectorized and ready to be utilized in the following layers.

crucial for feature extraction, yielding “feature maps” as the output. To mitigate information loss at the borders during convolution with a specific kernel size, “padding” is introduced. Padding adds zero values around the input, indirectly adjusting its size. The “stride” parameter controls the convolution density, where a larger stride results in less dense coverage. Post-convolution, these feature maps may contain an abundance of features, potentially leading to overfitting. “Pooling” serves to reduce this redundancy by condensing the feature maps. The complete workflow of a 2-dimensional CNN system is depicted in Figure 3.5.

### 3.2.3 Wavelet Neuron Network Model

While CNNs have found widespread application in insurance and actuarial science due to their proficiency in handling structured data, Wavelet Neural Networks (WNNs) offer unique advantages for flood risk management. This section provides a detailed discussion on the advantages of WNNs and explains the rationale for employing them to analyze the fluvial component within our proposed GHDL framework. We begin with an introduction to the Discrete Wavelet Transform (DWT), which forms the foundation of the WNN approach.

The DWT is a linear operation that decomposes a signal in the space of square-integrable functions over the real numbers,  $L^2(\mathbb{R})$ , into a series of segments, facilitating a more efficient and accurate representation of signals. Here,  $\mathbb{R}$  represents the set of real numbers. In the

context of DWT, wavelets are defined as families of functions  $h_{ab}$ ,

$$h_{ab}(x) = |a|^{-1/2} h\left(\frac{t-b}{a}\right), \quad a, b \in \mathbb{R}, a \neq 0, \quad (3.2)$$

generated from a single base function  $h$  through the processes of dilation and translation (Daubechies, 1988). This method allows for the analysis of various signal components at different scales, making DWT a powerful tool in signal processing.

Select a function  $\phi$  from  $L^2(\mathbb{R})$  such that its family of translations, denoted as  $\{\phi(x - k), k \in \mathbb{Z}\}$ , forms an orthonormal set. Here,  $\mathbb{Z}$  represents the set of whole numbers. The wavelet family associated with  $\phi$  is defined as follows:

$$\phi_{jk} = 2^{j/2} \phi(2^j t - k), \quad j, k \in \mathbb{Z}, \quad (3.3)$$

It is evident that  $\phi_{jk}$  can be reformulated to match the format of Equation (3.2) by setting  $a = 2^{-j}$  and  $b = 2^j k$ . Utilizing  $\phi_{jk}$  as base functions, we can construct a series of linear spaces  $\{V_j \subset L^2(\mathbb{R}), j \in \mathbb{Z}\}$ , where

$$V_j = \left\{ \sum_k A_k \phi_{jk}(x) : \sum_k |A_k|^2 < \infty \right\}, \quad j \in \mathbb{Z}.$$

Assuming that  $\phi$  is selected such that the spaces are nested,

$$V_j \subset V_{j+1}, \quad j \in \mathbb{Z}, \quad (3.4)$$

and that

$$\bigcup_{j \in \mathbb{Z}} V_j \text{ is dense in } L^2(\mathbb{R}), \quad (3.5)$$

we then define the complementary set of  $V_j$  relative to  $V_{j+1}$ , creating another sequence of linear spaces:

$$W_j = V_{j+1} - V_j. \quad (3.6)$$

There exist a function  $\psi$ , whose family of translations is orthonormal. The wavelet space of

$\psi$ , expressed as

$$\psi_{jk} = 2^{j/2} \psi(2^j t - k), \quad j, k \in \mathbb{Z}, \quad (3.7)$$

linearly spans the sequence of spaces  $\{W_j \subset L^2(\mathbb{R}), j \in \mathbb{Z}\}$ , where

$$W_j = \left\{ \sum_k D_k \psi_{jk}(x) : \sum_k |D_k|^2 < \infty \right\}, \quad j \in \mathbb{Z}.$$

If a function  $\phi$  is chosen in accordance with the conditions specified in Equations (3.4) and (3.5), any function  $g \in L^2(\mathbb{R})$  can be uniquely represented as a convergent series within  $L^2(\mathbb{R})$ , with respect to  $\phi$  and  $\psi$  (see [Härdle et al., 1998](#), for details of proofs and the methodology for deriving  $\psi$  once  $\phi$  is selected.):

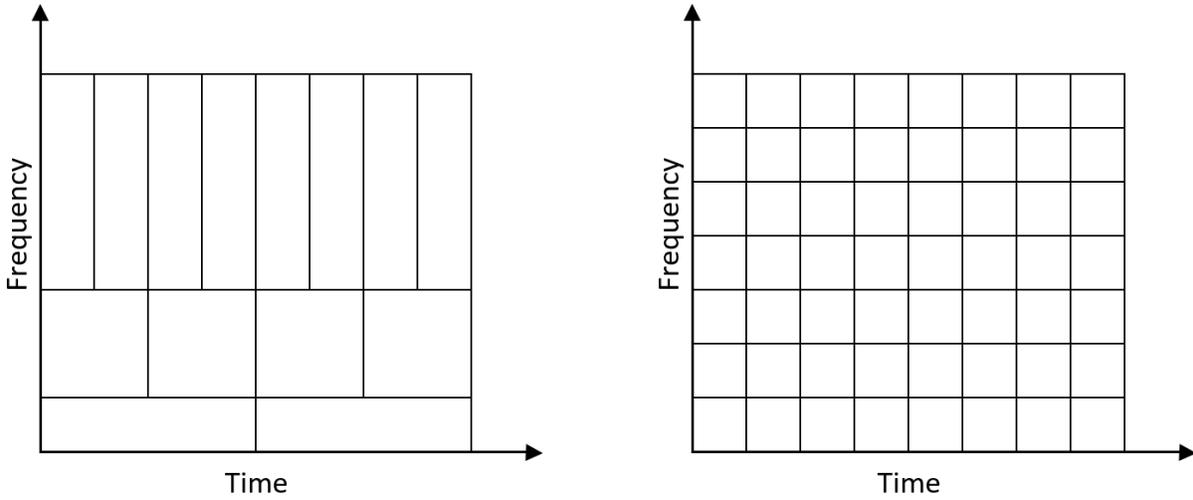
$$g(x) = \sum_k A_k \phi_{0k}(x) + \sum_{j=0}^{\infty} \sum_k D_{jk} \cdot \psi_{jk}(x). \quad (3.8)$$

This relation is known as the discrete wavelet transform (DWT) of  $g$ . The selected function  $\phi$  is called scale wavelet, and the derivation of  $\psi$  is called mother wavelet.

The expansion in Equation (3.8) exhibits the property of localization in both time and frequency domains. The summation over  $k$  corresponds to localization in time (shifts of functions  $\phi_{j0}(x)$  and  $\psi_{jk}(x)$ ). Conversely, summation over  $j$  corresponds to localization in frequency. As  $j$  increases, the associated frequency of  $\psi_{jk}(x)$  becomes higher, allowing for the analysis of finer details in the frequency domain.

Given this nature of representation, the coefficients  $D_{jk}$  are referred to as “detail coefficients.” These coefficients capture the high-frequency information at each level of the decomposition, pinpointing the more nuanced aspects of the signal. Conversely, the coefficients  $A_k$ , known as “approximation coefficients,” represent the remaining signal information that is not captured by the detail coefficients. They essentially provide a smoothed or averaged version of the signal, reflecting its broader trends.

The time and frequency localization properties of the DWT endow it with several desired characteristics, making it a powerful tool for analyzing river flow data in flood risk management. First, river flow data is typically non-stationary, influenced by systemic factors such



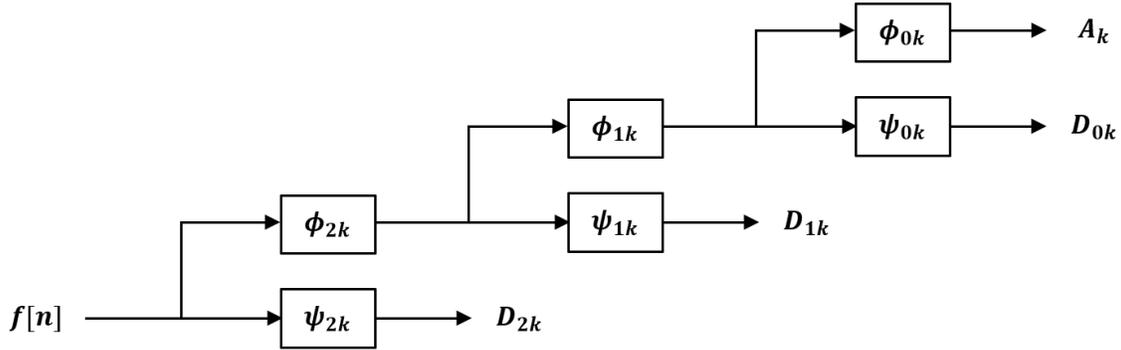
(a) Discrete wavelet transform

(b) Fast Fourier transform

**Figure 3.6.** Coverage of the time-frequency plane of the wavelet and Fourier based transforms

as climate change and seasonality, as well as occasional events like heavy rainfall. The ability of DWT to localize both time and frequency allows for a more precise analysis of transient features and anomalies in the signal (Daubechies, 1988). Second, DWT facilitates multi-resolution analysis of river flow data. This enables the decomposition of data into different scales, effectively capturing long-term trends, such as seasonal variations, and short-term events, such as flash floods, within a unified framework. Third, river flow data can often be noisy, affected by factors like sensor inaccuracies or environmental interference. DWTs are particularly useful in denoising this data, thereby enhancing the quality of information crucial for making informed predictions and decisions. Owing to these capabilities, DWTs have been extensively applied in flood risk management to transform and analyze various data types (see Shafaei and Kisi, 2016; Seo et al., 2015; Kumar et al., 2015, among others).

In practical applications, when dealing with a discrete signal  $g[n]$  and a scaling function  $\phi : \mathbb{R} \rightarrow \mathbb{R}$ , the wavelet coefficients  $A_k$  and  $D_{jk}$  are computed by progressively projecting the signal onto the scaling and mother wavelet functions. This projection starts from the highest selected frequency and proceeds to the lowest selected frequency (see Figure 3.7 for



**Figure 3.7.** A 3-level discrete wavelet transform system

an example):

$$A_k = \sum_n g[n] \phi_{0k}(n), \quad (3.9)$$

$$D_{jk} = \sum_n g[n] \psi_{jk}(n). \quad (3.10)$$

Notably, filtering a low-frequency signal using a higher-frequency filter does not result in the loss of information. Consequently, the original signal  $g[n]$  can always be reconstructed accurately using the acquired wavelet coefficients.

WNNs synergize the DWT with neural networks (NNs) by inputting decomposed wavelet coefficients into a fully connected neural network architecture. Recognized as one of the state-of-the-art machine learning methodologies in flood risk management (Mosavi et al., 2018), WNNs have been shown to enhance model accuracy, particularly in comparison with traditional Fourier transform methods (Shafaei and Kisi, 2016).

### 3.3 Performance in Flood Risk Prediction

In this section, we will assess the prediction performance of the proposed GHDL model. Meteorological and hydraulic information will be fed into the GHDL model for forecasting flood risk. The output will then be utilized to inform insurance pricing. This “two-step”

procedure is more advantageous, for two reasons. First, flood and weather have better data availability, compared to policy and claim data. Notably, while flood risk and related data have been globally accessible since the early or mid-20th century, thanks largely to satellite technology, policy and claim data for specific regions may only span a few years. Second, meteorological and hydraulic data, which are often publicly available, have better transparency. In contrast, the insurance loss data are typically available to insurers only. This improvement in flood risk assessment, in turn, refines flood insurance pricing by enabling the model to more accurately trace the underlying flood risks.

To empirically test the proposed GHDL structure, we selected a cluster of six cities located in the Mississippi River basin. This basin drains an area of about 3.2 million km<sup>2</sup> and encompasses 41% of the 48 contiguous states of the U.S. The chosen cities are Indianapolis, Nashville, Memphis, Louisville, Cincinnati, and St. Louis, all with a significant history of inland flooding and a high population exposed to flood risks. Our area of interest is displayed in Figure 3.8, and the connectivity of the six selected cities in the GHDL models is as displayed in Figure 3.9. We have excluded the coastal city of New Orleans from our study due to the sparse distribution of gauging stations in the lower Mississippi River, a limitation also noted by Quinn et al. (2019). Additionally, we have omitted other major cities that are far from our area of interest or have limited records of historical flood loss, such as Chicago, Des Moines, and Oklahoma City.

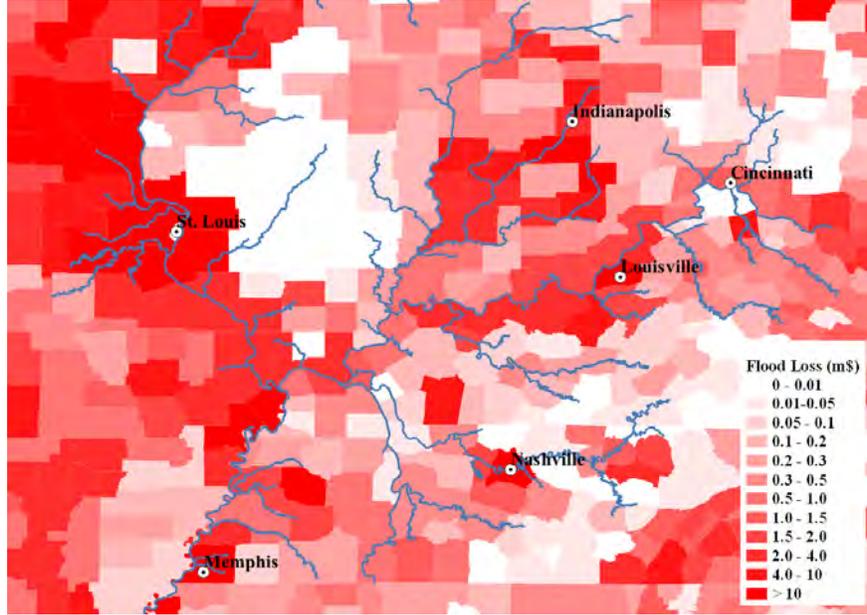
### 3.3.1 Data

#### River Flow Data

We obtained daily river flow data from 1990 to 2020 from the United States Geological Survey (USGS).<sup>3</sup> The selection of gauging stations for each city, as detailed in Table 3.1, is based on the criteria of data availability and proximity to the city’s centroid. To enhance the interpretability of extreme values in our analysis, we apply a logarithmic transformation to the river flow data. Subsequently, this logarithmically transformed data is normalized to

---

<sup>3</sup>The data can be accessed at <https://waterdata.usgs.gov/nwis/sw>.



**Figure 3.8.** The cluster of six cities the Mississippi river basin

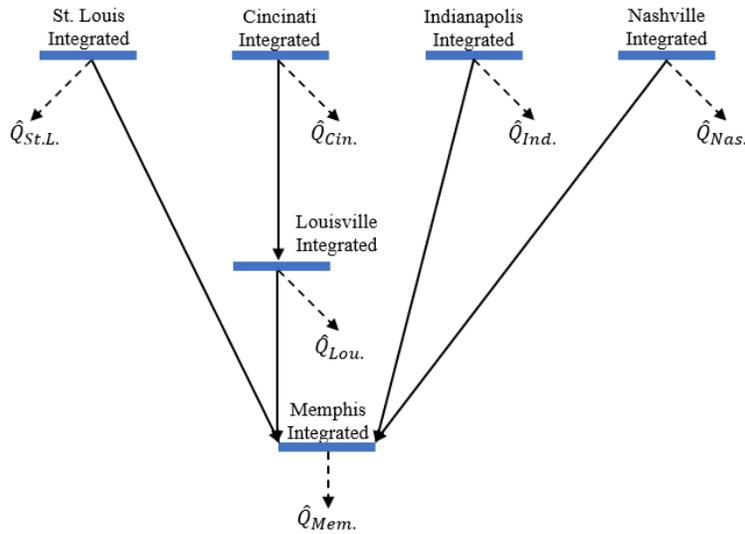
This figure plots the area of interest of this paper. A cluster of six cities located within the Mississippi River basin is selected. The gradient color in the map represents the historical flood losses for each county in the area.

fall within a range of 0 to 1 for each station, to facilitate model training. The distribution of the transformed river flow data, segregated by station, is shown in Figure 3.10. Given the length of river within the area of interest, we feed the previous seven days of river flow data into the fluvial segments of the GHDL structure.

**TABLE 3.1**  
**Selected Cities and their Corresponding Counties and Stations**

This table presents the information of six selected cities in the area of interest, including the population of the city in year 2021 rounded to one thousand, the corresponding county names and Federal Information Processing Standard (FIPS), and the corresponding gauging station names and id.

City	Population	County Name	FIPS	Station ID	Station Name
Indianapolis	869,000	Marion, IN	18097	03352500	Fall Creek at Millersville
Nashville	693,000	Davidson, TN	47037	03434500	Harpeth River Near Kingston Springs
Memphis	651,000	Shelby, TN	47157	07031650	Wolf River at Germantown
Louisville	618,000	Jefferson, KY	21111	03302000	Pond Creek Near Louisville
Cincinnati	323,000	Hamilton, OH	39061	03245500	Little Miami River at Milford
St. Louis	305,000	St. Louis, MO	29510	05587000	Macoupin Creek Near Kane



**Figure 3.9.** The GHDL structure

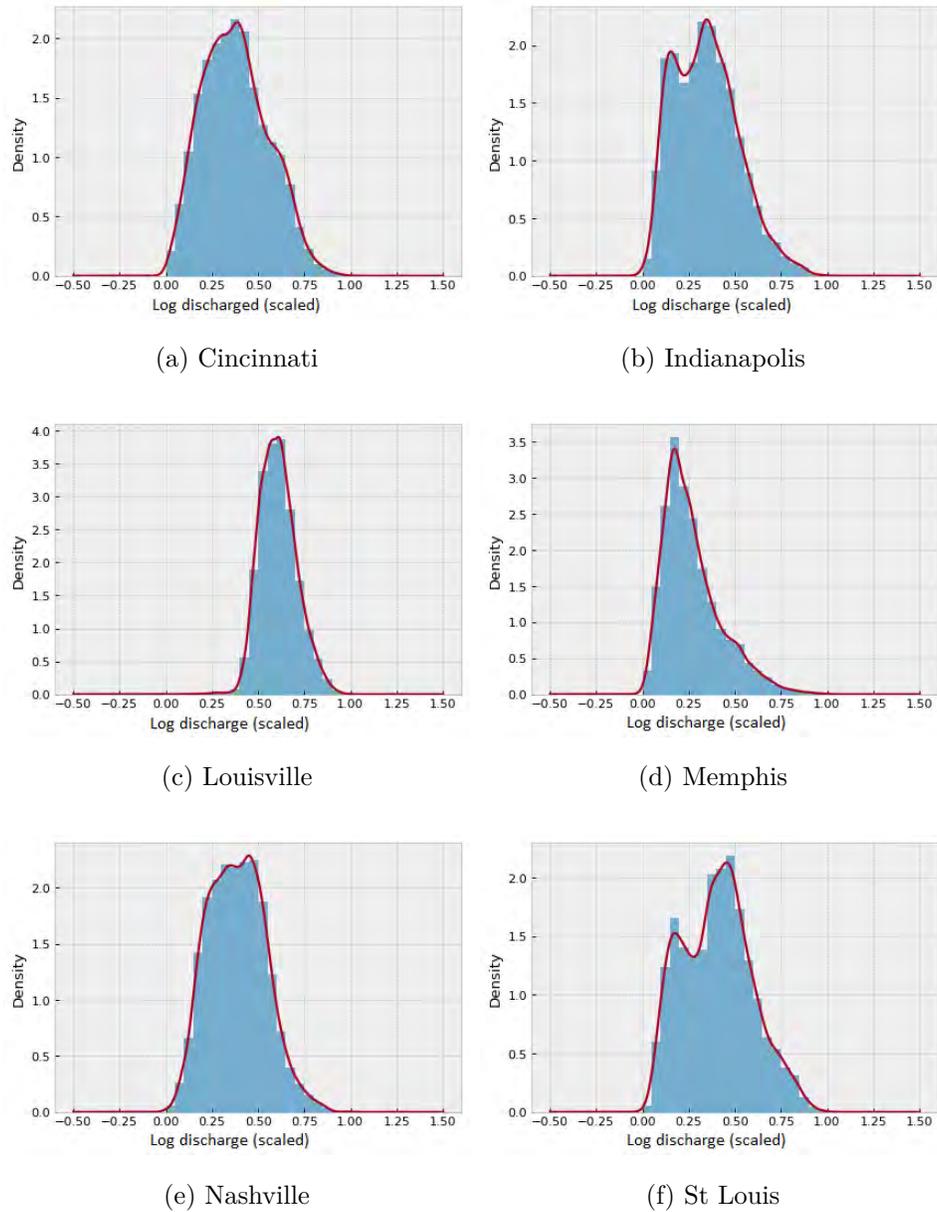
This figure displays the connectivity of the six selected cities in the area of interest in the GHDL models.

## Precipitation Data

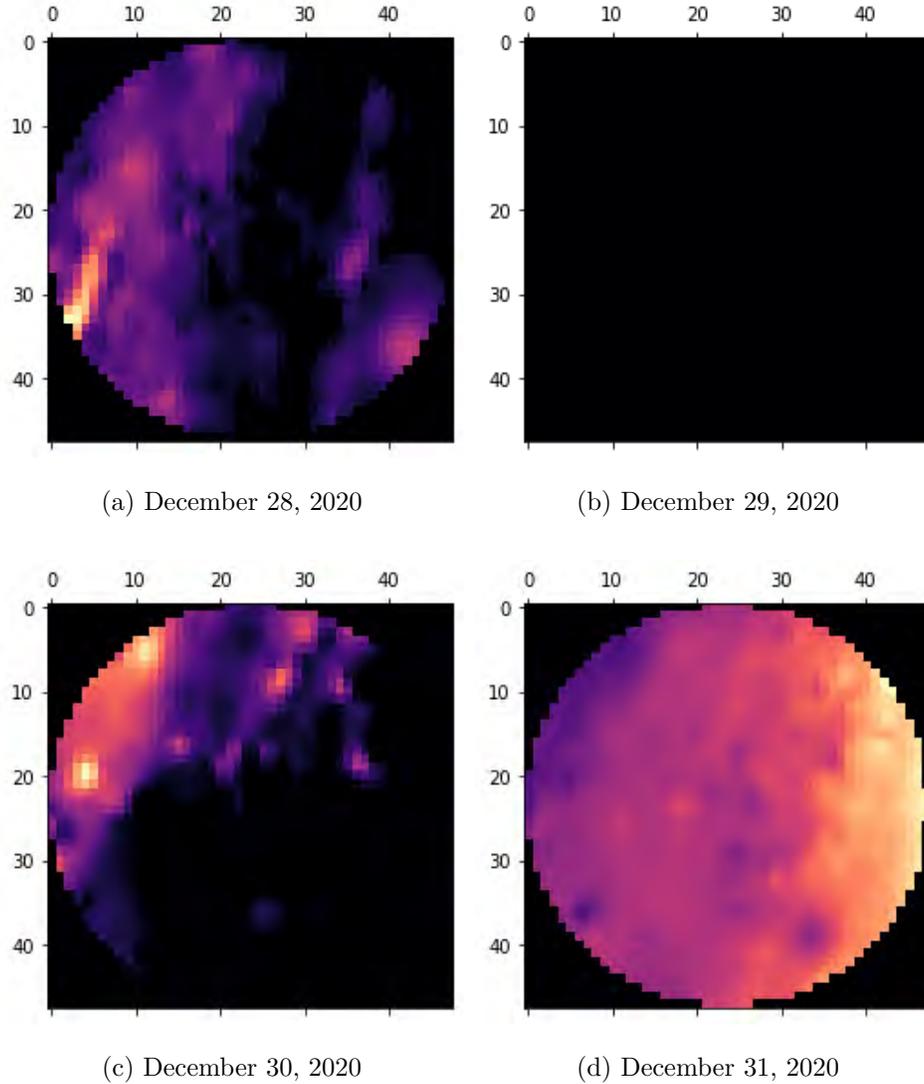
As outlined in Section 3.2.1, our proposed GHDL framework utilizes image precipitation data within its pluvial segments to evaluate flood risks associated with heavy rainfall.

The daily image precipitation data for the period 1990 to 2020 are sourced from the PRISM Climate Group, organized in  $4 \text{ km} \times 4 \text{ km}$  grid formats.<sup>4</sup> For each city under study, we generate 100 km radius circular buffers around the city centroids to extract precipitation data of a consistent size for input into the CNNs. This radius is arbitrarily selected to be large enough to cover each city and a significant portion of the surrounding area. Consequently, for each city, we obtain a daily time series of precipitation image data with a resolution of  $50 \times 50$  pixels ( $50 = 100 \times 2/4$ ), which is then fed into the CNNs. Figure 3.11 displays samples of daily precipitation rasters for Indianapolis covering the period from December 28 to 31, 2020. In these rasters, lighter colors indicate higher precipitation levels, while darker colors indicate lower precipitation levels, bounded at zero.

<sup>4</sup>The PRISM image precipitation data is accessible at <https://prism.oregonstate.edu/>.



**Figure 3.10.** The distribution of the transformed river flow data by station



**Figure 3.11.** The precipitation raster of Indianapolis from December 28 to December 31 2020

### 3.3.2 Performance Metrics

To measure the prediction performance of the proposed GHDL model, we consider two accuracy metrics. The first one is the relative root mean square error (RRMSE) of flood risk. We define flood risk using extreme flood events, specifically considering river flow levels that exceed their Value at Risk (VaR) at the  $\alpha \in (0, 1)$  level.

More specifically, let  $F_{Q_i}$  denotes the distribution of  $Q_{it}$ , then  $VaR_\alpha(Q_i) = F_{Q_i}^{-1}(1 - \alpha)$ . For each level of exceedance probability  $\alpha = \{10\%, 5\%, 1\%, 0.5\%, 0.25\%\}$ , we define the

observed flood indicator,  $\mathbb{1}_{Q_{it}}(\alpha)$ ,

$$\mathbb{1}_{Q_{it}}(\alpha) = \begin{cases} 1 & \text{if } Q_{it} \geq VaR_{\alpha}(Q_i), \\ 0 & \text{otherwise.} \end{cases} \quad (3.11)$$

Similarly, the predicted flood indicator  $\mathbb{1}_{\hat{Q}_{it}}(\alpha)$  is defined as:

$$\mathbb{1}_{\hat{Q}_{it}}(\alpha) = \begin{cases} 1 & \text{if } \hat{Q}_{it} \geq VaR_{\alpha}(Q_i), \\ 0 & \text{otherwise.} \end{cases} \quad (3.12)$$

The RRMSE for exceedance probability  $\alpha$  tracks the relative prediction error between the observed flood risk  $Q_{it}$  and the predicted flood risk  $\hat{Q}_{it}$ , conditional on the occurrence of a flood event as indicated by  $\mathbb{1}_{Q_{it}}(\alpha) = 1$ . That is:

$$RMSE(\alpha) = \sqrt{\frac{1}{\alpha \cdot IT} \sum_{i,t} \left( \frac{Q_{it} - \hat{Q}_{it}}{Q_{it}} \right)^2 \mathbb{1}_{Q_{it}}(\alpha)}. \quad (3.13)$$

The second metric employed to assess the predictive performance of our GHDL model is recall, a statistical measure commonly used in machine learning and information retrieval to evaluate the accuracy of a model. Recall represents the fraction of correctly predicted positive instances out of all actual positive instances. In the context of flood risk prediction, it measures the fraction of flood events that were correctly identified, where both observed and predicted indicators agree that a flood event occurred ( $\mathbb{1}_{Q_{it}}(\alpha) = \mathbb{1}_{\hat{Q}_{it}}(\alpha) = 1$ ):

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}. \quad (3.14)$$

In practical terms, recall is a critical metric for insurers, as it reflects the model's ability to accurately identify potential loss occurrences due to flood events.

We compute the performance metrics at the city level and then obtain a population-weighted average to aggregate them. This approach ensures a balanced representation that reflects both large and small populations within the studied areas.

### 3.3.3 Model Selection for Flood Risk Prediction

As mentioned before, hydraulic data and meteorological data typically have a longer sample period than policy and claim data, which is commonly the case for insurers in practice. We first determine the optimal structure of the GHDL model that provides the most accurate prediction for flood risk using historical hydraulic and meteorological information, and then test its pricing performance.

We split the data into a 70% training set and a 30% test set, preserving the temporal order of our data. We train the GHDL models on the training set for each set of model structures and evaluate the prediction accuracy of each model using the test set. To improve the speed of convergence, we adopt a momentum method. Additionally, to prevent overfitting, we add the  $L^2$  norm of the parameters to the loss function to regularize the training.

We compare models with different choices of hyperparameters built within the GHDL structure. Table 3.2 showcases the results from these models, each with distinct layer structures. The columns in Table 3.2 detail different WNN layer structures for the fluvial segments and the rows represent various CNN layer structures for the pluvial segments, where “ $i - j$ ” indicates a two-hidden-layer system with  $i$  and  $j$  neurons in the first and second hidden layers, respectively. For CNN layers, we standardize the convolutional kernel size at 3 and the stride and padding at 1 across all models, in line with the input size of the image precipitation data.<sup>5</sup> Furthermore, max pooling with a kernel size and padding of 2 is employed in all models to minimize redundancy.

It is notable that the GHDL structure consisting of two hidden layers in the fluvial segment and two hidden layers in the pluvial segment achieves the highest prediction accuracy, particularly in the most extreme scenarios tested, namely when  $\alpha$  is 0.5% or 0.25%. Therefore, we consider this structure as the baseline and analyze its performance later in flood risk forecasting and pricing in the following sections.

---

<sup>5</sup>Considering our 50-by-50 image precipitation data matrices, this setup maintains their spatial resolution, ensuring detailed feature capture without downsizing. This configuration adheres to the “CNN Output Size Formula.”

**TABLE 3.2**

**Geo-Hierarchical Deep Learning Model Selection for Flood Risk Forecasting**

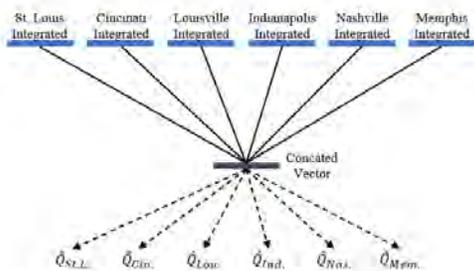
This table presents the results obtained from deep learning models with different layer structures. The columns present different WNN layer structures for the fluvial segments of the models and the rows represent different CNN layer structures for the pluvial segments of the models, where “ $i - j$ ” indicates a two-hidden-layer with  $i$  and  $j$  neurons in the first and second hidden layers, respectively. For each model, we calculate training and test RRMSE and Recall for exceedance probabilities  $\alpha = \{10\%, 5\%, 1\%, 0.5\%, 0.25\%\}$ .

		Fluvial Segments							
		1 hidden layer 64				2 hidden layer 128-64			
Pluvial Segments	Exceedance Probability	RMSE		Recall		RMSE		Recall	
		Training	Test	Training	Test	Training	Test	Training	Test
2 hidden layers 16-32	10.00%	18.65%	23.65%	77.87%	75.18%	14.14%	16.66%	81.08%	78.55%
	5.00%	16.81%	21.81%	72.89%	69.78%	9.94%	12.66%	86.72%	83.26%
	1.00%	17.51%	22.51%	70.02%	66.23%	7.05%	12.53%	90.45%	82.76%
	0.50%	17.14%	22.14%	68.35%	64.99%	5.42%	13.26%	92.28%	79.54%
	0.25%	22.53%	27.53%	67.27%	60.84%	6.05%	11.03%	94.07%	86.94%
3 hidden layers 16-32-64	10.00%	21.67%	26.67%	78.47%	75.54%	26.09%	31.09%	85.27%	82.03%
	5.00%	19.50%	24.50%	72.62%	68.27%	24.43%	29.43%	82.33%	78.50%
	1.00%	18.46%	23.46%	65.99%	61.22%	21.15%	26.15%	84.51%	78.84%
	0.50%	18.89%	23.89%	67.28%	62.13%	19.97%	24.97%	83.72%	79.22%
	0.25%	23.75%	28.75%	66.96%	59.09%	24.16%	29.16%	86.25%	79.17%
4 hidden layers 16-32-64-128	10.00%	27.97%	32.97%	76.30%	72.60%	26.47%	31.47%	85.37%	81.63%
	5.00%	23.70%	28.70%	70.21%	65.46%	21.51%	26.51%	81.88%	77.41%
	1.00%	19.10%	24.10%	66.21%	61.20%	16.81%	21.81%	80.32%	74.33%
	0.50%	16.19%	21.19%	64.58%	60.07%	15.34%	20.34%	78.55%	72.51%
	0.25%	21.35%	26.35%	65.46%	58.11%	21.26%	26.26%	80.64%	73.32%

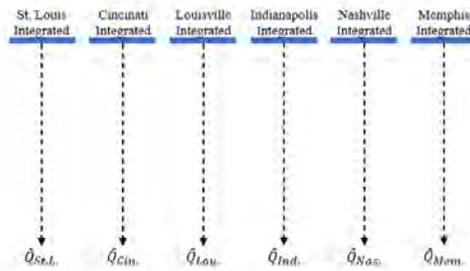
### 3.3.4 GHDL Performance in Flood Risk Forecasting

To evaluate the performance in flood risk forecasting of the proposed GHDL model, we employ two benchmark models that utilize the same pluvial and fluvial segments to integrate local hydraulic and meteorological data, as displayed in Figure 3.12.

- The Spatial-Flat Deep Learning Model (SFDL Model): This model leverages the same spatial and temporal data as our proposed GHDL model. However, it differs in its approach to spatial data integration. The SFDL Model does not account for the geographical connectivity among cities. Instead, it relies on a non-hierarchical, flat concatenation of data vectors from all cities for river flow predictions, disregarding the geographical relationships between these urban areas.
- The Local-Context Deep Learning Model (LCDL Model): Contrary to the SFDL and GHDL models, the LCDL Model exclusively utilizes locally integrated information for training. This model focuses on processing data within a constrained local context, omitting broader spatial interconnections.



(a) Spatial-Flat Deep Learning Model



(b) Local-Context Deep Learning Model

**Figure 3.12.** The model architecture of two benchmark structures on the area of interest

In addition, we also include a WNN-based machine learning model as an additional benchmark. This model, trained exclusively on the river flow dataset, is included due to its recognition as a state-of-the-art approach in current literature.

The geological hierarchy naturally divides the cities into two categories: those with cities upstream (descendant cities) and those without (leaf cities). For simplicity, we will use these terms going forward. As illustrated in Figure 3.9, there are two descendant cities (Louisville and Memphis) and four leaf cities (Indianapolis, Cincinnati, Nashville, and St. Louis) in our area of interest. Our GHDL models allow descendant cities to receive direct information from their upstream cities; hence, we expect to see greater improvements in flood risk forecasting for descendant cities than for leaf cities.

The performance of the GHDL structure in flood risk forecasting is presented in Table 3.3. Panels A and B present the results for descendant cities and leaf cities, respectively. As expected, the GHDL model brings more significant improvements to the descendant cities in flood risk forecasting than its benchmarks, compare to leaf cities. This distinction supports the superiority of our model, as it demonstrates that the hierarchical structure effectively leverages more informative data pathways.

Panel A demonstrates that the GHDL model outperforms its benchmarks in the descendant cities and most notably in the most extreme test cases. Specifically, the GHDL model achieves an RRMSE of 8.74% on the testing set when focusing on the most extreme 0.25% flood events. This performance surpasses that of the benchmark models, which exhibit RRMSE values ranging approximately between 21% and 25%. This trend of improvement is also evident in the recall metric, with GHDL model attaining a high recall rate of 91.26% in the test sample, while benchmarks only manage recall rates between 75% and 79%.

It is also noteworthy that in moderate risk scenarios, the SFDL model shows comparable capabilities to the GHDL model. For instance, with an exceedance probability of 10%, the SFDL model's performance (17.37% RRMSE and 82.63% recall) closely approaches that of the GHDL model (14.17% RRMSE and 85.83% recall). However, this parity does not extend to more extreme risk scenarios, where the SFDL model falls behind.

Panel B reveals that both the GHDL and LCDL models outperform the WNN and SFDL models in leaf cities. Notably, the LCDL model emerges as the top performer in the test sets for leaf cities, with the GHDL model ranking second (difference ranges from 1.8% to 4.4% for RRMSE). This result is not surprising given the relatively simpler task assigned to

**TABLE 3.3**  
**Performance in Flood Risk Forecasting**

This table details the performance of flood risk factors generated by various deep learning frameworks, as evaluated in both training and testing samples. Performance metrics include the Relative Root Mean Squared Error (RRMSE) and Recall. Panels A and B separately report these performance measures for descendant cities and leaf cities, respectively.

Panel A: Cities with Upstream Nodes (Descendant Cities)									
Measure	Exceedance Probability	WNN Model		SFDL Model		LCDL Model		GHDL Model	
		Training	Test	Training	Test	Training	Test	Training	Test
RMSE	10.00%	37.26%	38.12%	6.57%	17.37%	22.99%	25.50%	10.04%	14.17%
	5.00%	31.51%	30.46%	7.41%	19.07%	19.50%	21.45%	8.42%	10.52%
	1.00%	22.93%	28.23%	11.96%	22.40%	16.45%	22.10%	7.03%	11.66%
	0.50%	13.69%	29.52%	10.09%	29.91%	9.69%	25.39%	1.46%	12.27%
	0.25%	19.41%	25.17%	14.17%	25.22%	18.51%	21.25%	4.83%	8.74%
Recall	10.00%	62.74%	61.88%	93.43%	82.63%	77.01%	74.50%	89.96%	85.83%
	5.00%	68.49%	69.54%	92.59%	80.93%	80.50%	78.55%	91.58%	89.48%
	1.00%	77.07%	71.77%	88.04%	77.60%	83.55%	77.90%	92.97%	88.34%
	0.50%	86.31%	70.48%	89.91%	70.09%	90.31%	74.61%	98.54%	87.73%
	0.25%	80.59%	74.83%	85.83%	74.78%	81.49%	78.75%	95.17%	91.26%
Panel B: Cities without Upstream Nodes (Leaf Cities)									
Measure	Exceedance Probability	WNN Model		SFDL Model		LCDL Model		GHDL Model	
		Training	Testing	Training	Testing	Training	Testing	Training	Testing
RMSE	10.00%	20.92%	22.52%	12.69%	27.89%	15.85%	16.57%	16.77%	18.26%
	5.00%	14.85%	14.92%	13.12%	27.70%	9.90%	12.30%	10.91%	14.04%
	1.00%	8.96%	12.65%	11.98%	34.95%	7.38%	10.30%	7.07%	13.09%
	0.50%	10.83%	13.33%	15.96%	35.01%	6.89%	12.03%	7.96%	13.89%
	0.25%	9.97%	18.95%	14.78%	36.69%	6.71%	8.12%	6.83%	12.50%
Recall	10.00%	79.08%	77.48%	87.31%	72.11%	84.15%	83.43%	83.23%	81.74%
	5.00%	85.15%	85.08%	86.88%	72.30%	90.10%	87.70%	89.09%	85.96%
	1.00%	91.04%	87.35%	88.02%	65.05%	92.62%	89.70%	92.93%	86.91%
	0.50%	89.17%	86.67%	84.04%	64.99%	93.11%	87.97%	92.04%	86.11%
	0.25%	90.03%	81.05%	85.22%	63.31%	93.29%	91.88%	93.17%	87.50%

the LCDL model, which is dedicated to processing local precipitation and river flow data for a leaf city, tailored to make predictions exclusively for its designated location. In contrast, each corresponding submodel in the GHDL model must not only predict outcomes for the leaf city itself but also account for the downstream impact, thereby carrying a dual predictive responsibility. Consequently, with equivalent model complexity, the LCDL model inherently holds an advantage in leaf city predictions.

In practical applications, insurers could leverage river flow and precipitation data from upstream regions of these leaf cities, transforming them into descendant cities to enhance prediction accuracy. Moreover, simpler models tend to converge more effectively in scenarios where flood risk factors are less complex. Despite incorporating spatial information,

the SFDL model underperforms, even compared to the WNN benchmark. This could be attributed to the complexity of the model, potentially leading the optimization algorithm astray. A noticeable discrepancy between training and test results in the SFDL model further signifies this issue, indicating that the SFDL model may suffer from an overfitting.<sup>6</sup> This highlights the notion that indiscriminately feeding extensive climate data without appropriate constraints may not yield optimal results in flood risk management contexts.

In summary, our forecasting results show that the GHDL model is more competitive than its benchmarks in flood risk forecasting for all exceedance probabilities. The improvement is most significant in descendant cities, and the model partially addresses the issue of performance degradation that commonly associated with increasing model complexity.

## 3.4 Performance in Flood Insurance Pricing

In this section, we assess the pricing performance of the GHDL model using the policy and claim datasets from National Flood Insurance Program (NFIP).

### 3.4.1 National Flood Insurance Program

The NFIP was established by the U.S. government to provide insurance coverage for losses caused by flooding. Its main objective is to offer financial assistance to property owners who experience flood damage, which is typically not covered by standard homeowner’s insurance policies in the U.S. As of 2020, the program covers 3,053 out of the 3,143 counties in the U.S. The Federal Emergency Management Agency (FEMA) maintains the NFIP policy and claim datasets.<sup>7</sup> See [Kousky and Michel-Kerjan \(2017\)](#) and [Michel-Kerjan and Kousky \(2010\)](#) for detailed examinations of the NFIP’s claim and policy dataset.

To obtain the NFIP policy and claim datasets, we source the data from OpenFEMA and filter it by the Federal Information Processing Standards (FIPS) number of the selected

---

<sup>6</sup>We have repeated the training for the SFDL model for several times and confirmed the discrepancy is systematic.

<sup>7</sup>The NFIP policy and claim dataset is accessible at <https://www.fema.gov/about/openfema/data-sets>.

**TABLE 3.4**  
**Descriptive Statistics of the NFIP**

This table provides descriptive statistics for the NFIP policy and claim dataset from 2009 to 2020. Panel A details the policy dataset for each city, including the total number of effective policies (“Policies in-force”), the average coverage amount per policy (“Coverage per policy”), and the average premium collected per policy (“Premium per policy”). It also calculates “Flood-prone policies”, representing the proportion of policies in the Special Flood Hazard Area (SFHA) relative to total policies. NFIP defines SFHA as an area with a 1-percent or greater annual chance of flooding. Panel B summarizes the claim dataset by city, showing the total number of claims (“Claims”) and the average payment amount per claim (“Amount paid per policy”).

Panel A: Policy Dataset						
	Cincinnati	Indianapolis	Louisville	Memphis	Nashville	St. Louis
Policies inforce (k)	25.3	68.5	69.1	51.7	85.8	4.8
Coverage per policy (\$k)	210	187	170	271	234	296
Premium per policy (\$)	1227	1007	852	566	777	1589
Flood prone policies (%)	76.6	84.1	81.6	32.3	54.9	57.7
Panel B: Claim Dataset						
	Cincinnati	Indianapolis	Louisville	Memphis	Nashville	St. Louis
Claims (k)	1.9	2.2	5.7	1.1	4.5	0.7
Coverage per policy (\$k)	111	105	115	139	169	88
Amount paid per policy (\$k)	15.0	7.6	14.7	21.6	37.4	11.8
Flood prone policies (%)	92.6	92.2	95.0	70.7	83.4	92.3

cities’ respective counties. The policy dataset consists of 305,121 policy-level observations from 2009 to 2020, including information on policy effective and termination dates, premiums, coverage, deductibles, first policy dates, and house characteristics. The claim dataset comprises 16,001 transaction-level observations from 2009 to 2020. We focus on the policies that cover residential properties and match the policy and claim data by comparing the policy features and building characteristics of the two datasets. See Table 3.4 for summary statistics of the policy and claim data. For training and testing data segmentation, we designate January 1, 2017, as the dividing date within our 2009-2020 policy-claim dataset, resulting in a roughly 70%-30% training-test split that preserves temporal order.

### 3.4.2 Flood Insurance Pricing

Let  $Y_{kt}$  be a random variable representing the total flood loss associated with the  $k$ th insured building located near gauging site  $i$  during the time period  $[t - 1, t]$  ( $i = 1, \dots, I; k =$

$1, \dots, K, t = 1, \dots, T$ )<sup>8</sup>. In practice, we estimate  $Y_{kt}$  by multiplying its corresponding damage ratio  $U_{kt} \in [0, 1]$ , with the insured building value  $\theta_k$  that is,  $Y_{kt} = \theta_k \cdot U_{kt}$ . Given a deductible  $d_k$  and a limit  $l_k$  the total claim of a flood insurance policy  $Y_{kt}^L$  is defined as below:

$$Y_{kt}^L = \begin{cases} 0 & \text{if } Y_{kt} < d_k, \\ Y_{kt} - d_k & \text{if } d_k \leq Y_{kt} < l_k, \\ u_k - d_k & \text{if } Y_{kt} \geq l_k. \end{cases} \quad (3.15)$$

To simplify the analysis, we take  $d_k = 0, l_k = \theta_k$ , that is  $Y_{kt}^L = Y_{kt}$ . Conceptually, for a specific insured building  $k$ , there should exist a monotonically increasing damage function  $h_k : Q_{it} \rightarrow U_{kt}$  that maps the river flow observation  $\{Q_{it} = q_{it}\}$  to the corresponding loss event  $\{U_{kt} = u_{kt}\}$ . Let  $f_{Q_{it}}$  represent the probability density function of  $Q_{it}$ . The pure premium,  $p_{kt}$ , can be obtained by applying this damage function to the expected value of the river flow:

$$\begin{aligned} p_{kt} &= \mathbf{E} [Y_{kt}^L] = \int_0^\infty u_{kt} \theta_k \cdot f_{Q_{it}}(q_{it}) dq_{it} = \int_0^\infty h_k(q_{it}) \theta_k \cdot f_{Q_{it}}(q_{it}) dq_{it} \\ &= \theta_k \mathbf{E} [h_k(Q_{it})]. \end{aligned} \quad (3.16)$$

In this study, we utilize a set of generalized linear models (GLM) to establish the relationship between river flow, building characteristics, and the loss ratio. We denote the claim frequencies (i.e., number of claims incurred) as  $N_{kt}$ , and the severity (i.e., loss ratio per claim) as  $U_{kt}$ , then we have

$$\mathbf{E} [Y_{kt}^L] = \theta_k \mathbf{E} [N_{kt}] \mathbf{E} [U_{kt}]. \quad (3.17)$$

We create separate GLMs for  $N_{kt}$  and  $U_{kt}$ , utilizing the forecasted expectation of river

---

<sup>8</sup>Note that  $Y_{kt}$  does not include the indexing of  $i$  because the  $k$ th insured building corresponds to only one gauging site in our setting. Thus, indexing flood loss by  $k$  is sufficient to represent it.

flow,  $\widehat{Q}_{it}$ , provided by the GHDL model, and the individual-level building characteristics,  $\mathbf{x}_k = (x_{k1}, \dots, x_{kp})'$ :

$$\nu_{kt} = \mathbb{E} [N_{kt} | \widehat{Q}_{it}, \mathbf{x}_k] = g_n^{-1} (\lambda_n \widehat{Q}_{it} + \mathbf{x}'_k \beta_n), \quad (3.18)$$

$$\mu_{kt} = \mathbb{E} [U_{kt} | \widehat{Q}_{it}, \mathbf{x}_k] = g_u^{-1} (\lambda_u \widehat{Q}_{it} + \mathbf{x}'_k \beta_u), \quad (3.19)$$

where  $g_n$  and  $g_u$  are link functions with respect to frequency and severity GLM models, respectively;  $\lambda_n$  and  $\lambda_u$  are their regression coefficients which corresponds to flood risk factor; and  $\beta_n$  and  $\beta_u$  are  $p \times 1$  regression coefficients associated with building characteristics. Here,  $(\lambda_n, \beta_n)$  and  $(\lambda_u, \beta_u)$  are estimated from historical observations. Finally, the net premium can be calculated as follow:

$$p_{kt} = \theta_k \cdot \nu_{kt} \cdot \mu_{kt}. \quad (3.20)$$

### 3.4.3 GLM Model Summary

We test the pricing performance of GHDL generated flood risk factors on one-year term flood insurance policies, consistent with industry practice, noting that over 99% of policies in our sample have this term duration. For each year  $t$ , we use the optimal GHDL model structure identified in Section 3.3.3, extending the forecast period to 365 days and utilizing the prior year's  $(t - 1)$  meteorological and hydraulic data for daily river flow predictions. These predictions are aggregated on a monthly basis to form 12 risk factors.<sup>9</sup> Additionally, we include city dummy and loss month dummy variables, along with building characteristics, to refine our pricing model.

Table 3.5 presents the results of the GLMs integrating risk factors obtained from the GHDL model and benchmarks on the training set. Besides the two benchmark models (SFDL and LCDL) considered in Subsection 3.3.4, we also compare the pricing results with a GLM model that overlooks flood risk factors (Non-Flood Risk Factor GLM), to illustrate how flood risk factors generated by different DL models can improve the pricing. Panel A

---

<sup>9</sup>Our evaluation encompasses a range of aggregation methods, including monthly mean, median, and maximum, as well as the average of the top 10 and top 5 monthly risk factors. Our analysis of various aggregation methods reveals that the monthly mean yields the most precise pricing.

details the frequency GLMs, utilizing a log-link function and Poisson distributions for claim counts. Panel B, on the other hand, describes the severity GLMs, which employs a log-link function with Gamma distributions for claim amounts. We can see that the GHDL model is superior to all benchmarks in both frequency and severity GLMs in terms of goodness-of-fit, indicated by lower values in negative log-likelihood (NLL) and Akaike information criterion (AIC). Coefficients  $\lambda_1$  to  $\lambda_{12}$  in GLM models are mostly significant, indicating that the extracted risk factors from the GHDL models play a critical role in the pricing models. In Section 3.5.1, we will interpret the pricing model by providing a detailed analysis of the relative importance of the risk factors in the GLM models.

### 3.4.4 Net Premium and Solvency Capital Requirement

Figure 3.13 shows the net premium charged under different forecasting models alongside the actual losses incurred every year. The grey vertical line distinguishes between the in-sample and out-of-sample years. The light blue bars represent the actual claims paid, dark blue bars represent the net premium based on the proposed GHDL model, orange and red bars stand for the net premiums based on the benchmarks of the SFDL and LCDL models, respectively, and green bars show premiums from a model with no weather information incorporated. We have a few interesting observations from this figure. First, the GHDL model consistently provides the most accurate predictions for claims, evident in both in-sample and out-of-sample data. Second, incorporating spatial and hierarchical structures significantly enhances flood risk pricing; excluding these structures results in greater model variance and less consistency. As we can see, although the SFDL and LCDL models perform competitively in-sample, they exhibit significant prediction variability in out-of-sample years. Finally, weather information is critical in flood risk pricing. We can see that the net premium from the model with no weather information performs poorly in both samples.

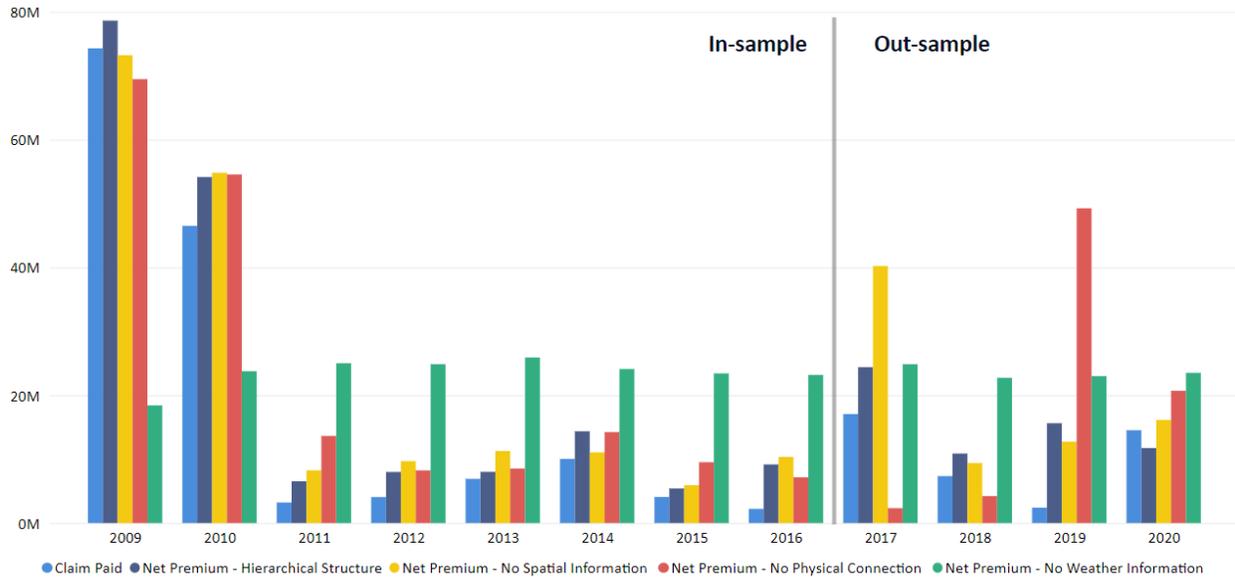
We report the pricing performance of various deep learning structures on the test sample in Table 3.6. To mitigate the impact of occasional outliers, we employ a bootstrapping approach, repeatedly sampling policy and claim pairs from the test sample 10,000 times. The net premium for each policy is predetermined as of December 31 of the preceding

**TABLE 3.5**

**General Linear Model Result in Flood Risk Pricing**

This table displays the results of generalized linear models in flood risk pricing, focusing on pricing power of flood risk factors generated by different deep learning models. Panel A shows the regression results for the frequency GLM, while Panel B addresses the severity GLM. Each panel presents coefficients for model-generated flood risk factors across months, from January to December (denoted as  $\lambda_1$  to  $\lambda_{12}$ ). “\*\*\*” indicates a  $p$ -value  $< 0.01$ , “\*\*” signifies a  $p$ -value  $< 0.05$ , and “\*” denotes a  $p$ -value  $< 0.1$ .

Panel A: Frequency GLM				
	Non-Fld Factor GLM	SFDL-based GLM	LCDL-based GLM	GHDL-based GLM
$\lambda_1$		-10.39***	-4.09***	-1.43**
$\lambda_2$		5.64***	5.26***	9.98***
$\lambda_3$		0.76	-1.12***	-8.56***
$\lambda_4$		-3.83***	-0.05	2.67***
$\lambda_5$		17.39***	8.32***	12.16***
$\lambda_6$		-20.28***	-6.57***	-9.06***
$\lambda_7$		10.83***	3.55***	5.34***
$\lambda_8$		6.56***	4.35***	7.12***
$\lambda_9$		-8.45***	4.09***	-7.98***
$\lambda_{10}$		-2.45**	-11.18***	-7.95***
$\lambda_{11}$		-0.48	-4.52***	0.93
$\lambda_{12}$		-2.07***	6.02***	1.48**
NLL	20754	18419	18079	17830
AIC	41573	36588	36247	35747
Building characteristics	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes
County fixed effect	Yes	Yes	Yes	Yes
Panel B: Severity GLM				
	Non-Fld Factor GLM	SFDL-based GLM	LCDL-based GLM	GHDL-based GLM
$\lambda_1$		1.09	-1.63**	2.45***
$\lambda_2$		-1.73*	0.26	0.35
$\lambda_3$		0.54	-0.57	-2.18***
$\lambda_4$		0.31	-0.05	2.09**
$\lambda_5$		1.94***	3.01***	0.94
$\lambda_6$		-2.82***	-4.22***	-3.67***
$\lambda_7$		3.80***	3.63*	3.75**
$\lambda_8$		-3.07***	1.27	3.6352
$\lambda_9$		4.09***	-4.48**	-4.12**
$\lambda_{10}$		-3.65***	1.65	-0.70
$\lambda_{11}$		0.64	-0.20	-0.12
$\lambda_{12}$		0.65	-1.80***	0.71
NLL	-1743	-1796	-1776	-1802
AIC	-3427	-3509	-3470	-3522
Building characteristics	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes
County fixed effect	Yes	Yes	Yes	Yes



**Figure 3.13.** Net premium vs. Loss

This figure plots the total net premium charged under different risk forecasting models and the actual losses incurred by year.

year. Subsequently, we calculate the aggregate total net premium and total claims for each calendar year. Panel A of the table presents the average annual total net premium results, while Panel B details the solvency capital requirements as computed under the Solvency II framework.<sup>10</sup>

The proposed GHDL structure leads to a 33.5% reduction in net premium relative to the pricing model that overlooks flood risk information. The SFDL and LCDL benchmarks also provide net premium improvements of 16.6% and 18.7%, respectively. Specifically, the proposed GHDL structure reduced the net premium by USD 7.87 million per annum. The lower net premium in our model results from the improved accuracy of the proposed GHDL flood model.

Next, we discuss the implications of our model on capital requirements. Under the Solvency II framework, insurers are required to hold a minimum level, known as the Solvency Capital Requirement (SCR), to cover their risks with a 99.5% probability over a one-year period (EIOPA, 2023). To simplify our discussion, we assume that the insurer operates ex-

<sup>10</sup>The predominant regulatory framework for insurance companies in the U.S. is the Risk-Based Capital (RBC) rather than Solvency II. However, we opt for the Solvency II framework in our calculations because the solvency capital requirement (SCR) under this framework is less restricted by company-specific information.

clusively within this flood insurance business line and has no other operational income. This assumption aligns the SCR directly with the Basic Solvency Capital Requirement (BSCR) for this specific insurance program.

BSCR is calculated as the difference between the total net premium charged, and the Value-at-Risk (VaR) at the 0.5% level ( $VaR_{0.5\%}$ ) on loss. The introduction of the proposed GHDL model results in a notable decrease in net premiums, which in turn slightly reduces the VaR on losses—from 0.05 million to -0.23 million. Importantly, the GHDL model reduces the BSCR by 32.3%, outperforming the benchmarks. For comparison, the SFDL and LCDL benchmarks provide BSCR improvements of 15.3% and 18.3%, respectively. This reduction in BSCR signifies a potential decrease in the capital required to be held by insurers, thereby reducing their capital costs and enhancing financial efficiency. The superior performance of the GHDL model demonstrates its effectiveness in more accurately predicting flood risks and optimizing financial resources under regulatory requirements.

**TABLE 3.6**  
**Performance in Flood Insurance Pricing**

This table compares the out-of-sample pricing performance of various deep learning frameworks. Panel A presents the improvements in net premium, while Panel B illustrates the enhancements in solvency capital requirements. We calculate the Basic Solvency Capital Requirement (BSCR) as the difference between difference between total net premium and the VaR at 0.5% level on loss. These comparisons shed light on the effectiveness of different deep learning approaches in insurance financial management.

Panel A: Net Premium Results				
	Non-Flood Risk Factor GLM	SFDL-based GLM	LCDL-based GLM	GHDL-based GLM
Panel A: Net Premium Results				
Net Premium (\$m, an.)	23.52	19.61	19.12	15.65
Net Premium Reduced (\$m, an.)		3.91	4.40	7.87
Net Premium Improvement (%)		16.62%	18.71%	33.45%
Panel B: Solvency Capital Requirement Results				
	Non-Flood Risk Factor GLM	SFDL-based GLM	LCDL-based GLM	GHDL-based GLM
Value-at-Risk 0.5% (\$m, an.)	0.05	-0.26	-0.04	-0.23
Expected Shortfall 0.5% (\$m, an.)	-0.41	-0.58	-0.42	-0.63
BSCR (\$m, an.)	23.47	19.87	19.17	15.88
BSCR Reduced (\$m, an.)		3.60	4.30	7.59
BSCR Improvement (% , an.)		15.34%	18.34%	32.33%

## 3.5 Interpretability of the GHDL Model

Previous sections have demonstrated the superiority of the GHDL model in flood risk forecasting and flood insurance pricing. It is also important to improve the interpretability and transparency for deep learning-based models such as GHDL (Chen et al., 2023b; Cong et al., 2019). In this section, we study the relative importance (see Section 3.5.1) and partial dependence (see Section 3.5.2) of the risk factors extracted from GHDL model that are used for pricing in GLM models.

In our interpretability analysis in this section, we separate the annual flood insurance policy to twelve monthly insurance policies, each priced by the forecasted risk factor of that month. It is important to note that we use the same flood risk forecasting results as in Section 3.4. Therefore, the structure of the monthly flood insurance program allows policyholders to select specific months they wish to cover, which provides flexibility based on anticipated risk.

To ensure that the forecasted risk factor reflect not just historical averages but also short-term weather risks, we incorporate a lagged flood risk factor into our GLM models. This factor considers the average monthly river flow from the previous year for each city, providing a dynamic adjustment to environment variations.

### 3.5.1 Relative Importance Analysis

Figure 3.14 shows the top 20 most relatively important factors in the GHDL-based frequency and severity GLM models. The relative importance of a factor is calculated as its corresponding regression coefficient ( $\lambda$  or  $\beta_p$  as in equation 3.4.2) times the correlation coefficient between the factor and dependent variable, scaled to 0 and 100. This figure shows that the GHDL forecasted risk factors significant relative importance than the building characteristics.

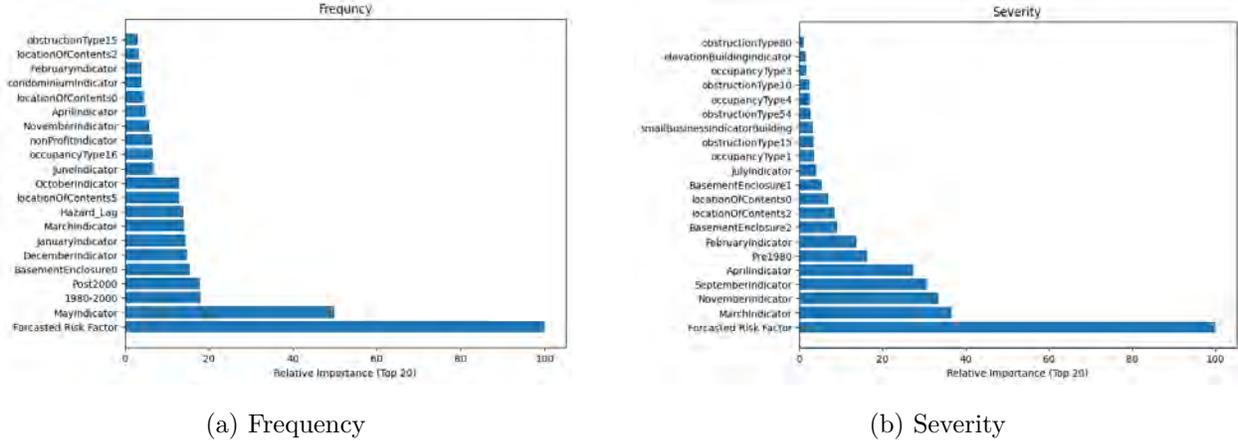


Figure 3.14. GHDL foretasted flood risk factor relative importance

This figure displays the top 20 most relatively important factors in the frequency and severity GLM models. The figure shows that the forecasted risk factors are more important than the building and housing characteristics.

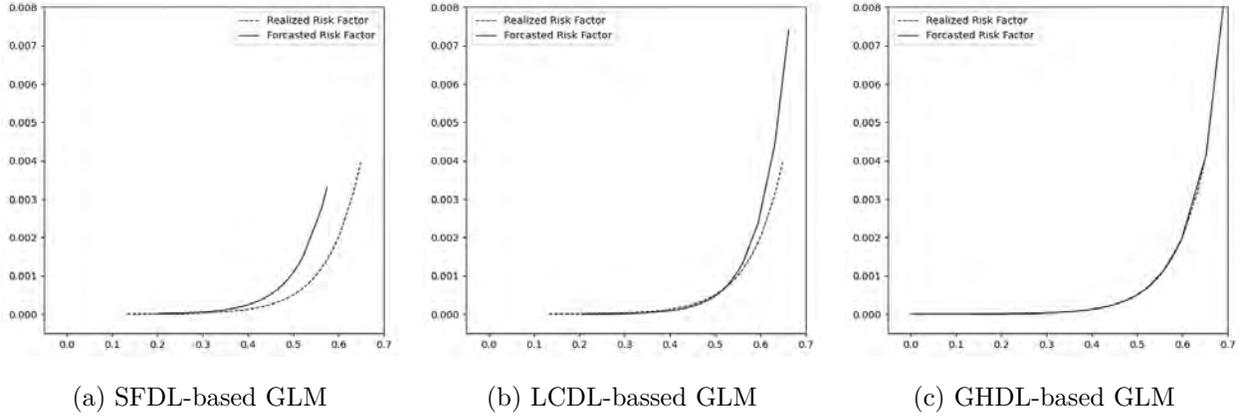
### 3.5.2 Partial Dependence Analysis

To further demonstrate the explanatory power of the forecasted risk factors for flood risk, we compute the partial dependence of the GHDL forecasted risk factor in the fitted frequency GLM model, as shown in Figure 3.15(c).<sup>11</sup> In addition, we also calculate the partial dependence of the forecasted risk factor from the GLMs fitted using risk factors forecasted by the SFDL and LCDL models, which are shown in Figure 3.15(a) and 3.15(b), respectively. For comparison, we fit a GLM model using realized river flow data, with its partial dependence plotted in dashed lines across all three figures. We can see that the proposed GHDL structure’s forecasted risk factors most closely approximate the anchor.

## 3.6 Conclusion Remarks

In this research, we introduce a GHDL framework specifically designed to enhance flood risk modeling through the effective integration of high-resolution meteorological and hy-

<sup>11</sup>We omit the partial dependence analysis for the severity GLM in our study, because we observed that the severity GLM shows a stronger association with the building characteristics variables and exhibits significant imbalances across these variables, which is consistent with the observations of Kousky and Michel-Kerjan (2017). Consequently, the partial dependence analysis would be less meaningful and informative for the severity GLM in this context.



**Figure 3.15.** Forecasted risk factor partial dependence

This figure displays the partial dependence between the risk factors produced by the GHDL model, the SFDL model, the LCDL model, and the claim counts. To provide a point of reference, we fit a GLM model that replace the forecasted flood risk factor with the realized river flow and plot its partial dependence in the three panels with dashed lines.

draulic data. The geographical hierarchical structure in this framework is tailored to reflect geographical locations, providing the adaptability needed to model complex flood scenarios. We apply the framework to data from the Mississippi River and show that the GHDL model is superior to traditional deep learning benchmarks that do not incorporate geographical connectivity and/or spatial information. This improved accuracy establishes a more robust foundation for flood risk assessment and insurance pricing.

Importantly, the risk factors generated by our model not only enhance net premiums but also contribute to reducing the solvency capital requirements. This is particularly beneficial for the National Flood Insurance Program. Additionally, our model employs a global optimization approach to effectively process spatial and temporal data, while its geo-hierarchical structure enhances model interpretability. The GHDL framework can potentially address complexities such as anthropogenic factors, promising to provide more informed and reliable strategies in flood risk management.

Future research can explore the application of the GHDL model to a broader range of geographical regions and river systems, especially in developing countries, to validate the versatility and scalability of the model. Additionally, incorporating different climate scenarios into the GHDL framework would be interesting. Exploring the model’s capacity to dynamically adapt to changing environmental factors amid climate uncertainty could further

refine its predictive capabilities and enhance its utility in real-world applications. We leave these discussions for future work.

# Chapter 4

## Bridging the Protection Gap: A Tax Redistribution Solution Under a Private-Public Partnership Framework

### 4.1 Introduction

The escalating impacts of climate change have disproportionately magnified the risks of catastrophes in regions highly vulnerable to climate variability (Aczel, 2022; Thomas et al., 2018; Levy and Patz, 2015; Hirabayashi et al., 2013; Holland and Bruyère, 2013; Mann and Emanuel, 2006; Westerling and Bryant, 2007). Although evidence suggests that individuals are inclined to mitigate climate risk through insurance (Hossain et al., 2022; Botzen et al., 2009), the increase in net premiums induced by climate change, coupled with heightened capital requirements posed by these risks, often results in insurance costs exceeding what many households are prepared to pay (Netusil et al., 2021; Charpentier, 2007). As 37% of Americans do not have \$400 in liquid funds for an emergency<sup>1</sup>, this economic disparity has

---

<sup>1</sup>The statistic is from Economic Well-Being of U.S. Households in 2022 report published by Board of Governors of The Federal Reserve System. Accessable at: <https://www.federalreserve.gov/publications/>

further widened the protection gap in areas already at risk, underscoring the urgent need for innovative solutions to address these challenges.

The escalating threat of climate-related catastrophes has led insurers to retreat from high-risk areas, reducing coverage [Capacity](#) and exacerbating the climate protection gap. This trend has made obtaining P&C insurance increasingly difficult and costly in high climate risk regions ([Isidore and Nilsen, 2023](#)). For example, in California, emerging wildfire risks have led to a significant reduction in the number of domestic P&C insurers operating in the state—from 112 in 2012 to 90 in 2022, marking an approximate 20% decrease, in contrast to the 7% nationwide decline in P&C domestic insurers.<sup>2</sup> Several national insurers have withdrawn from the state ([Gall, 2023](#); [Morris and Botros, 2023](#)). Similar patterns have been observed in Florida ([Gall, 2023](#); [Hudak, 2022](#); [Isidore, 2023](#); [Leefeldt, 2024](#); [Rahman, 2023](#); [The Guardian, 2023](#)) and Louisiana ([Henderson, 2023](#)), indicating a widespread issue across various regions. Such trend underscores the pressing need for innovative solutions to address this growing challenge

Insurers prefer certain markets based on climate exposure, which is rational from a portfolio optimization point of view. However, this preference increases insurance supply in moderate-risk regions while expanding the protection gap in high-risk areas.

On the demand side, although the risk profile of regions with high climate exposure is deteriorating as climate change progresses, the increase in demand for insurance may not keep pace with the escalating risks ([Browne and Hoyt, 2000](#); [Andor et al., 2020](#); [Raschky and Weckhannemann, 2007](#)). The concept of the Samaritan’s dilemma, as described by [Buchanan \(1975\)](#), elucidates how governments in modern welfare states often find themselves implicitly compelled to provide public disaster relief (PDR) to those affected. This expectation of government disaster relief can logically diminish individual households’ incentives to secure insurance coverage, a phenomenon referred to as “charity hazard”. This dynamic poses a significant challenge to closing the protection gap from demand side in regions most vulnerable to climate change.

The charity hazard and its impact on the demand for climate insurance have been exten-

---

[files/2022-report-economic-well-being-us-households-202305.pdf](#)

<sup>2</sup>Data sourced from the NAIC P&C Market Share Report (2013-2022).

sively analyzed within the framework of the climate risk protection gap, both theoretically (Raschky and Weckhannemann, 2007) and empirically (Browne and Hoyt, 2000; Andor et al., 2020; Tesselaar et al., 2022). Raschky and Weckhannemann (2007) analyze the economic impact of governmental financial relief on the natural disaster insurance market, suggesting that charity hazard deters individuals from purchasing insurance and hampers the market's development by blocking the widespread adoption of natural hazard insurance. Browne and Hoyt (2000) examine the financial records of the United States' National Flood Insurance Program (NFIP) from 1983 to 1993, revealing that income and price significantly influence flood insurance purchasing decisions, with state-level purchases strongly correlated with the previous year's flood losses. Andor et al. (2020) examine survey data from German homeowners and find significant charity hazard among residents in flood-prone areas and a positive correlation between government aid and non-financial protections. Tesselaar et al. (2022) employ a partial equilibrium model to analyze the impact of charity hazard on the flood insurance protection gap across EU countries through 2050, finding that uncertainty in government compensation and increased flood risks reduce charity hazard. However, existing literature often overlooks a critical aspect: the funding mechanism of PDR plans. Like most public expenditures, these plans are typically funded through taxation, whether existing or imposed after the fact, in many countries. In nations with varied climate dynamics within their borders, the financial burden of taxation-funded PDR efforts in high-risk regions does not fall solely on the residents of these areas but is also shared by those in moderate-risk regions. Therefore, even without direct exposure, residents of moderate-risk regions are still impacted by liabilities induced by climate change.

The liability associated with PDR illustrates the negative externalities caused by inadequate coverage of climate risks. This means that populations indirectly affected by climate risks can also suffer the consequences of damages not covered in areas of high exposure. Evidence shows that climate risks can impact regions without direct exposure through economic linkages, such as disruptions in the economic supply chain (See, e.g., Er Kara et al., 2020; Ghadge et al., 2019; Pankratz and Schiller, 2022) and impacts on financial markets (Venturini, 2022). Furthermore, Ge (2021) finds that insurance companies with diversified portfolios in both life and P&C lines may increase their premiums for life insurance policies

following significant natural catastrophe losses in their property lines. This phenomenon highlights another way in which climate risk externalities are transferred from affected to unaffected individuals. Consequently, strategies to narrow the protection gap in high-risk areas could benefit residents across all risk regions.

A natural response to the withdrawal of private sector insurers from high-risk areas is the introduction of public insurance programs. [Charpentier and Le Maux \(2014\)](#) argue that public insurance offers a better solution for hedging against climate risk compared to private insurance because public entities can levy taxes ex-post to cover premiums when faced with deficits. This approach is considered sustainable under the assumption that governments, unlike private insurers, cannot go bankrupt. Public insurance is also cheaper than private insurance. However, relying solely on public insurance has its drawbacks. Issues of fairness arise, particularly in how premiums are settled through taxation. Additionally, public insurance pricing often reflects factors beyond actuarial calculations and introduces inefficiencies. For instance, the NFIP has been criticized for politically motivated subsidies that disproportionately benefit high-risk regions, often under the guise of protecting property values ([Michel-Kerjan and Kunreuther, 2011](#)). Similarly, Citizens Insurance in Florida, designed as an insurer of last resort for properties unable to secure insurance elsewhere, faces solvency challenges that necessitate capital injections to remain operational ([Born et al., 2021](#)).

The public-private partnership (PPP) framework is a promising alternative to pure private and public insurance programs for hedging against climate risks. PPP in insurance represents a collaborative framework where both the private and public sectors share risks and rewards in providing insurance coverage. In this model, the private sector's role is primarily focused on underwriting risks and managing insurance operations, leveraging its expertise in risk assessment and customer service ([Kunreuther, 2015](#)). On the other hand, the public sector often assumes a regulatory role, providing a safety net or backstop that ensures the sustainability and affordability of insurance, especially in high-risk areas. This partnership allows for a more efficient distribution of risk and can enhance the accessibility of insurance to a broader segment of the population. A notable example of a climate risk hedging PPP in action is the United Kingdom's Flood Re scheme. Flood Re is designed

to make flood insurance more affordable and accessible to households at the highest risk of flooding. Under this scheme, insurers pass on the flood risk associated with high-risk homes to Flood Re, which is funded by a levy on insurers. This approach ensures that homeowners in flood-prone areas can obtain insurance at a reasonable cost, demonstrating a successful application of PPP in mitigating climate risk through insurance.<sup>3</sup>

The Flood Re program in the UK and the NFIP in the US both aim to provide flood insurance coverage but employ distinct operational models. Flood Re operates as a private-public partnership, where the private sector handles the delivery of insurance, supported by public funding to mitigate financial risks. In contrast, the NFIP directly provides insurance to property owners facing flood risks. This program assumes full responsibility for underwriting and claims, with the government centrally managing all aspects of risk and insurance coverage.

While the PPP framework limits public involvement in market development and helps reduce the cost of acquiring insurance, it is not without its challenges. For instance, in the Flood Re program, as the government assumes responsibility for high-risk homes, the risk profile of insurance companies shifts. This transfer can lead to less prudent behavior among insurers, as they are relieved of the burden of managing their riskiest policies.

In this chapter, we propose a tax-redistribution solution to narrow the substantial protection gap in high-risk regions through a self-financed tax-redistribution scheme under the PPP framework. A tax will be levied on the surplus of insurance coverage in moderate-risk regions, with the proceeds used to augment coverage in high-risk areas. The collected taxes will be utilized through a dual approach that includes offering subsidies to high-risk regions and expanding coverage capacity via co-insurance. This strategy aims to effectively redistribute resources, ensuring that the heightened risk in vulnerable areas is mitigated.

We solve the optimal redistribution by modeling the rational demand and supply dynamics between residents and insurers. In our model, residents are characterized as [risk averse](#), [price takers](#), and [expected utility maximizers](#), while insurers are also characterized as [risk neutral](#), price takers, and expected utility maximizers. We differentiate climate risk

---

<sup>3</sup>Notably, Flood Re also receives criticisms for being economically inefficient and failing to be a vehicle to promote behavioral change in the interests of sustainable flood risk management ([Penning-Rowsell, 2015](#)).

by modeling it as a distinct distribution, separate from that of moderate risks. Additionally, PDR is represented within our model as a fixed proportion of the uncovered loss, activated based on the magnitude of economic loss.

Our model explains several phenomena. First, on the demand side, we observe that wealth transfer from moderate-risk to high-risk regions via public disaster relief diminishes demand in high-risk areas. Conversely, the liability shift from high-risk to moderate-risk regions amplifies demand among risk-prudent residents in moderate-risk region. Second, from the supply perspective, the increased demand for solvency capital compels insurers seeking a specific rate of return to offer less actuarially favorable premiums to high-risk regions.

By aligning the tax redistribution mechanism with insurers' supply dynamics, we identify an optimal tax redistribution plan. Given a tax collection amount that is exogenously determined, we derive an implicit solution for the proportions of co-insurance and premium subsidy. Most importantly, our analysis reveals that by balancing the trade-off associated with PDR liability, the tax redistribution plan can facilitate a Pareto improvement across the entire population, benefiting residents of all risk regions.

To our knowledge, this study is the first to explore how extreme climate risk liability is transferred to the indirectly exposed populations, in the literature of rational demand for climate insurance.

This project contributes to the ongoing debate regarding the government's role in climate risk protection. One body of literature advocates for the benefits of public insurance in climate risk hedging, highlighting its immunity to insolvency risk and its capacity to finance claims through ex-post funding ([Charpentier and Le Maux, 2014](#); [Bruggeman et al., 2010](#)). Conversely, another segment of the literature critiques the efficiency of public involvement in this domain ([Cummins, 2006](#); [Michel-Kerjan and Kunreuther, 2011](#); [Born et al., 2021](#)). In light of these perspectives, the PPP framework emerges as a promising alternative, offering a middle ground between the two viewpoints. This chapter builds upon the concept of private-public risk-sharing, guiding the redistribution of social resources to achieve maximization of social welfare.

The remainder of this chapter is organized as follows: Section 4.2 introduces the proposed tax redistribution framework. Section 4.3 characterizes the rational demand and supply dynamics within this framework, detailing the conditions for optimal redistribution. Section 4.4 concludes the chapter.

## 4.2 The Tax Redistribution Model

Consider two regions where one is exposed to high climate risks and the other only experiences moderate risks. We call the former the high-risk region (HR) and the latter the moderate-risk region (MR). To simplify our discussion, we assume the residents of the two regions are homogeneous in all aspects except for risk exposure. The risks of the residents in the two regions are denoted as  $\epsilon_H$  and  $\epsilon_M$ , with their means and standard deviations as  $\mu_H, \mu_M$  and  $\sigma_H, \sigma_M$ , respectively. It follows that  $\mu_H > \mu_M$  and  $\sigma_H > \sigma_M$ .

Let  $S_H(\cdot), S_M(\cdot)$  and  $D_H(\cdot), D_M(\cdot)$  denote the supply and demand function of insurance at HR and MR in the absence of tax redistribution, respectively. We denote  $Q_H^*$  and  $Q_M^*$  as the equilibrium coverage ratios in HR and MR, and denote  $\pi_H^*$  and  $\pi_M^*$  as the equilibrium premium in HR and MR, respectively. By definition,  $Q_H^*$  and  $Q_M^*$  solve the problem of  $S_H(Q) = D_H(Q)$  and  $S_M(Q) = D_M(Q)$ , with  $\pi_H = S_H(Q_H^*) = D_H(Q_H^*)$  and  $\pi_M = S_M(Q_M^*) = D_M(Q_M^*)$ .

We use the elasticity ( $E = (dQ/Q)/(d\pi/\pi)$ ) to measure the supply and demand change in response to premium change. Denote  $E_{SH}(Q)$  and  $E_{SM}(Q)$  as the elasticity of the insurers' supply curve in HR and MR at coverage ratio  $Q$ , and denote  $E_{DH}(Q)$  and  $E_{DM}(Q)$  as the elasticity of the residents' demand curve in HR and MR at coverage ratio  $Q$ .

With the presence of tax redistribution, a tax  $\tau$  will be charged to insurers operating in MR, which will then be utilized to subsidize premium or provide capital injection to insurers operating in HR.

### 4.2.1 Public Disaster Relief Liability

One reason residents may be reluctant to purchase insurance for catastrophe exposure is the expectation that the government will provide PDR following extreme catastrophic events. To model the payment of PDR, we assume that when  $\epsilon_H$  surpasses a certain threshold  $L$  at time  $t = 1$ , the government will initiate PDR to address the uncovered losses, financed through ex-post taxation, which is common practice. We assume that only residents in HR areas face extreme climate risks; therefore, PDR will never be triggered in MR. Nevertheless, our model remains robust if the assumption is modified to allow  $\epsilon_M > L$  to also trigger PDR, though with a lower probability, i.e.,  $P(\epsilon_M > L) < P(\epsilon_H > L)$ .

Although PDR is typically funded by contributions from both HR and MR regions, only residents of HR directly benefit from it. Consequently, the PDR scheme essentially constitutes a conditional wealth transfer from MR to HR. We define the parameters of public disaster relief as follows:

$$\text{Public Disaster Relief} = \begin{cases} \theta & \text{if } \epsilon_H \geq L, \\ 0 & \text{if } 0 \leq \epsilon_H < L, \end{cases} \quad (4.1)$$

where  $\theta$  is the wealth transfer from MR to HR, which is effectively the present value of future taxation collected from MR to finance the PDR. Apparently, PDR is pecuniarily unfair for the residents of MR. Therefore, in principle, it should be employed sparingly and only serve as a safety net.

### 4.2.2 Insurers' Incentive

For an insurer with a liability portfolio  $\omega = (Q_H, Q_M)'$ , where  $Q_H \geq 0$  and  $Q_M \geq 0$  denote the total coverage she provides to HR and MR separately, her insurance outcome at  $t = 1$  is represented as

$$R(\omega) = Q_H\pi_H + Q_M\pi_M - Q_H\epsilon_H - Q_M\epsilon_M. \quad (4.2)$$

Assume the government (planner) requires the insurer to set aside a solvency capital  $v_0(\omega)$  at  $t = 0$  to ensure her expected probability of solvency at  $t = 1$  is no less than a ratio  $\alpha^4$  i.e.

$$P(R(\omega) + v_1(\omega) \geq 0) \geq \alpha, \quad (4.3)$$

where  $v_1(\omega)$  represents the value of reserving  $v_0(\omega)$  at  $t = 1$ .

To focus our analysis on the liability aspect of insurers, we simplify their asset management by imposing two constraints. First, insurers cannot finance through borrowing. Therefore, their initial financial position,  $v_0(\omega)$ , together with any insolvency premium, cannot exceed their initial endowment, denoted as  $v$ . Although credit financing is uncommon among insurers due to the increased risk of insolvency it presents, our framework can still be extended to scenarios where credit financing is permissible. Second, we require insurers to invest their reserves exclusively in [risk-free assets](#), meaning  $v_1(\omega) = (1 + r_f) \cdot v_0(\omega)$ , where  $r_f$  represents the [risk-free rate of return](#). This assumption allows us to bypass the variability in investment outcomes.

If the insurer is solvent at  $t = 1$ , she can enjoy a wealth of  $R(\omega) + v_1(\omega)$ . If the insurer becomes insolvent at  $t = 1$ , she will file for bankruptcy. To mitigate this risk, insurers are required to pay an insolvency premium,  $\pi_i$ , to the planner at  $t = 0$ , ensuring coverage of their liabilities in case of bankruptcy. It is assumed that the planner is risk-neutral and that the insolvency premium is priced to be actuarially fair, as described below:

$$\pi_i = -E\{[R(\omega) + v_1(\omega)] \wedge 0\}. \quad (4.4)$$

Should an insurer become insolvent, the planner utilizes the insolvency premium, pooled from insurers, to address the remaining liabilities. Given the assumption that the planner cannot go bankrupt, the policies of residents are effectively guaranteed.

---

<sup>4</sup>Under Solvency II, the required solvency ratio is 99.5%

In sum, assuming risk neutral, the insurer's objective function is represented as

$$\begin{aligned} \max_{\omega, v_0} \quad & (1 + \beta)(v - v_0 - \pi_i) + E\{[R(\omega) + v_1(\omega)] \vee 0\}, \\ \text{s.t.} \quad & 0 \leq v_0 + \pi_i \leq v. \end{aligned} \tag{4.5}$$

Here,  $\beta$  represent the insurers **required rate of return**. Insurers generally prefer to offer coverage in MR rather than HR. This preference is influenced by two main factors. First, the higher average loss ( $\mu_H$ ) and variability of loss ( $\sigma_H$ ) in HR areas reduce the marginal utility for insurers of increasing coverage, unless this is compensated by significantly higher premiums. Second, the greater exposure to severe climate risks in HR areas poses challenges to the insurers' capacity. Thus, one may expect that the insurers' supply is more elastic to premium in MR than in LR, which we will formally discuss in corollary 4.2.

### 4.2.3 Residents' Incentive

We assume that residents' utility is solely influenced by their pecuniary outcomes. As discussed in Section 4.2.1, the wealth transfer from MR to HR regions via the PDR scheme can be represented by  $\theta(1 - Q_H)1\{\epsilon_H > L\}$ , consumed exclusively by HR but financed through ex-post taxation from the entire population. Consequently, the PDR scheme results in MR residents bearing a portion of the liabilities of HR residents.

Following the convention, we model the insurers' risk transfer as residents will cede a portion of their risks to the insurer, denoted as  $Q_H$  for HR and  $Q_M$  for MR. We describe the incentives for residents in both regions as follows. For residents in HR,

$$v_H = EU[w - Q_H\pi_H - (1 - Q_H)\epsilon_H + \theta(1 - Q_H)1\{\epsilon_H \geq L\}], \tag{4.6}$$

and for residents in MR,

$$v_M = EU[w - Q_M\pi_M - (1 - Q_M)\epsilon_M - \theta(1 - Q_H^*)1\{\epsilon_H \geq L\}]. \tag{4.7}$$

Here,  $w$  represents the initial wealth of the residents, and  $U(\cdot)$  is a von-Neumann Morgenstern

utility function that captures their risk aversion.

Notably, the utility of residents in MR partially depends on the equilibrium coverage outcome in HR ( $Q_H^*$ ), which is exogenous to them. Through taxing insurance offerings in MR to redistribute to HR, MR residents' coverage is reduced, creating a trade-off between local risk liability and PDR liability.

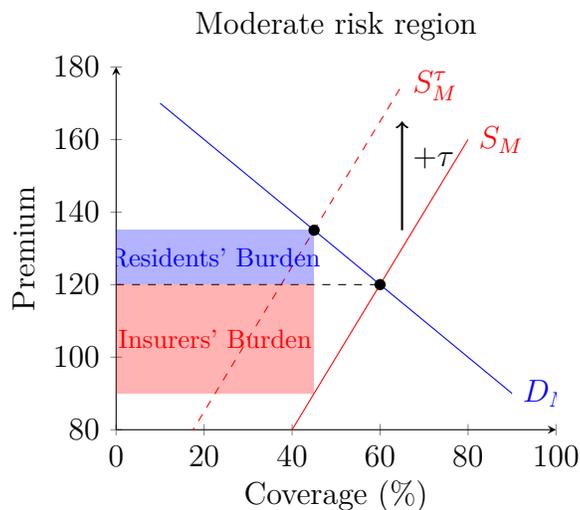
#### 4.2.4 Redistribution

As established in Section 4.2.3, the liability associated with PDR for MR residents can be mitigated by increasing the coverage ratio in HR areas. In this section, we propose a redistribution scheme that levies taxes on insurance policies in MR regions to either subsidize or provide the capital to enhance the coverage ratio in HR areas. The primary goal of this redistribution effort is to elevate the overall welfare of all residents.

#### Tax Collection

Tax collection from both insurers and residents can fulfill identical objectives. The former leads to an upward shift in the supply curve, while the latter results in a downward shift in the demand curve. We elucidate this concept by examining the scenario in which taxes are imposed on insurers. Let  $\tau$  represent the tax levied per policy issued in the MR region, and let  $Q_M^\tau$  denote the new equilibrium coverage ratio. The effect of tax collection on the equilibrium coverage ratio is illustrated in Figure 4.1, demonstrating the dynamic adjustments within the market.

Notably, the tax levied will increase the equilibrium premium for residents in MR. This adjustment will alter the insurance demand among MR residents. However, it remains uncertain whether this change in demand will be positive or negative. In Section 4.3.2, we will explore how the direction of this demand shift is influenced by the specifics of the residents' utility functions and considerations of premium fairness. Nonetheless, the variation in demand elasticity should be regarded as a matter of secondary importance.



**Figure 4.1.** Taxation on per policy written in low risk regions

In Figure 4.1, we depict the supply and demand curves as having a linear relationship between the premium and coverage ratio for illustrative purposes. However, our model does not necessitate this linear relationship.

### Subsidy Distribution

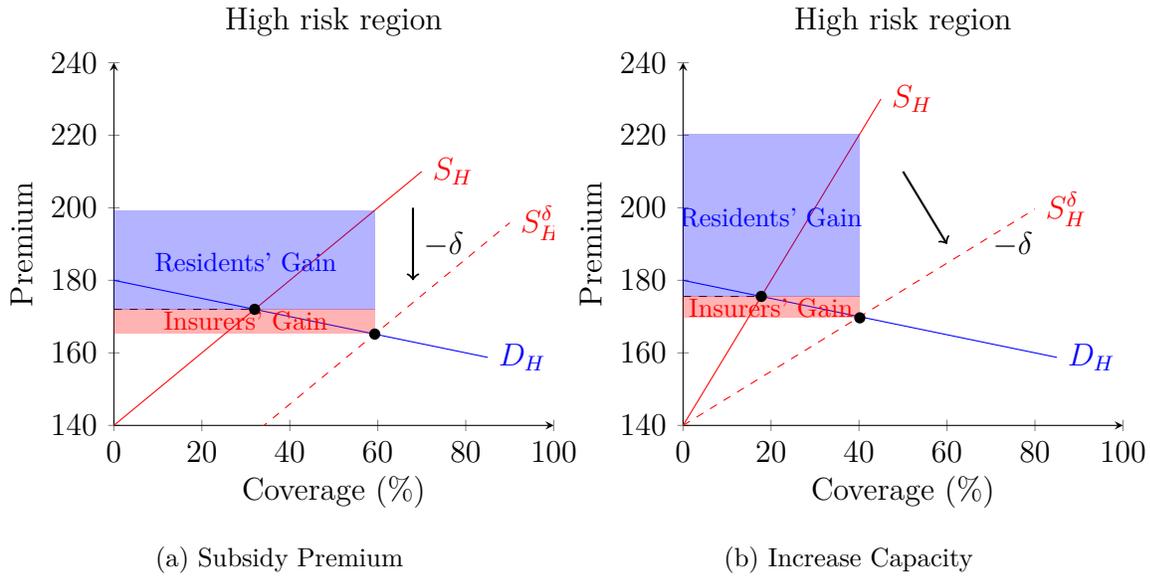
The optimal allocation of collected taxes is determined by the total available capacity within the HR insurance supply market. The supply decisions of insurers ( $\omega$ ) are influenced not only by risk-return efficiency but also by constraints related to their initial endowments (Cummins and Weiss, 2000; Cagle and Harrington, 1995), as shown in Equation (4.3). The endowment constraint dictates an insurer’s capacity, and collectively, the capacity of all insurers shapes the market’s capacity. This, in turn, is manifested through the elasticity of the supply curve. We will further discuss the impact of supply capacity on equilibrium coverage ratio in Section 4.3.3.

There are two ways the tax can be used to increase coverage supply in HR:

- **Subsidy Premium:** A fixed ratio of the premium will be subsidized by the scheme. When the collected taxes are redistributed to the insurers operating in HR as a premium subsidy per policy written, the insurers are able to supply a same amount of coverage with a lower premium charged from residents. Therefore, tax redistribution

through premium subsidy is effectively shifting the supply curve downward, as depicted in Figure 4.2(a).

- **Increase Capacity:** There are several ways to increase the insurance market capacity. Here, we propose that the planner provides proportional co-insurance with the insurer for the claims incurred in HR. Subsidizing the capital requirement essentially reduces the inelasticity of the supply curve, as depicted in Figure 4.2(b).



**Figure 4.2.** Description of tax redistribution in high and low risk regions

Offering co-insurance is not a sole strategy planners can employ to augment the market's capacity. For instance, planners could offer excess-loss insurance on top of the insolvency insurance for policies written in HR at no cost, utilizing tax revenues as a form of risk capital. When compared to the provision of excess-loss insurance, the approach of providing co-insurance carries several advantages. Considering that insurers can operate across both HR and MR, providing excess-loss insurance can alter the risk profile of the insurers, potentially leading to a scenario where capacities intended for HR are partially absorbed by the insurers' business in MR, detracting from the original objective. In addition, providing excess-loss insurance at no cost will alter the insurers' risk decisions. In our analysis, we advocate for the use of co-insurance as a planner's intervention to enhance market capacity.

When the elasticity of supply is low, subsidizing premiums would not be effective, given

the total capacity of the market restricts the insurers' ability to insure more risk in HR. On the other hand, when the elasticity of supply is high, meaning the capacity of the market is sufficient, providing further capacity would be less efficient. In specific market situations, the most efficient subsidy distribution method lies between these extremes.

As redistribution increases the optimal coverage ratio in HR areas, the protection gap narrows, reducing the tax burden of PDR for MR residents in the event of a catastrophe.

As both insurers and residents exhibit more elastic supply and demand functions in HR compared to MR regions<sup>5</sup>, the distribution scheme predominantly benefits property owners in HR, with the majority of the financial impact being absorbed by insurers in MR. Despite the fact that the tax imposition initially compromises the benefits for property owners, the resulting increase in the equilibrium coverage ratio within HR effectively reduces the post-tax liability of MR residents funding the PDR scheme. Therefore, the levy placed on MR areas holds the potential to facilitate a Pareto improvement for property owners in MR.

## 4.2.5 Planer's Problem

The planner's objectives can vary, each leading to a different problem formulation. In this section, we outline two potential objectives a planner might consider: first, achieving Pareto efficiency among the residents; and second, striving for Kaldor-Hicks efficiency. These objectives frame the planner's approach to optimizing the welfare of the residents through different efficiency criteria.

For the sake of discussion, denote  $\tau$  as tax collection at MR and  $\delta$  as tax redistribution at HR. With the tax redistribution scheme  $(\tau, \delta)$ , denote  $Q_H^\delta$  and  $Q_M^\tau$  as the equilibrium coverage ratio in both region,  $\pi_H^\delta = S_H^\delta(Q_H^\delta) = D_H^\delta(Q_H^\delta)$  and  $\pi_M^\tau = S_M^\delta(Q_M^\tau) = D_M^\tau(Q_M^\tau)$  as the equilibrium premium in both region. Therefore, under  $(\tau, \delta)$ , the HR residents' utility is

$$v_H^\delta = EU_H\{w_H - Q_H^\delta \pi_H^\delta - (1 - Q_H^\delta)\epsilon_H + \theta(1 - Q_H^\delta)1\{\epsilon_H > \theta\}\}, \quad (4.8)$$

---

<sup>5</sup>Later we will show in Corollary 4.2 and Corollary 4.1.

and the LR residents' utility is

$$v_M^\tau = EU_M\{w_M - Q_M^\tau \pi_M^\tau - (1 - Q_M^\tau)\epsilon_M - \theta(1 - Q_H^{\delta*})1\{\epsilon_H > \theta\}\}. \quad (4.9)$$

### Pareto Efficiency among Residents

To achieve Pareto efficiency, the planner must ensure that the collected tax is utilized in the most efficient way possible, specifically through the allocation of premium and capacity subsidies. They must then limit the tax collection to a point where the reduction in public aid liability is at least as great as the tax burden on residents. Therefore, the planner's objective involves a two-step maximization process:

$$\begin{aligned} \max_{\tau} \left( \max_{\delta} \left( v_H^\delta + v_M^\tau \right) \right), \\ \text{s.t. } v_H \leq v_H^\tau, \quad v_M \leq v_M^\tau, \quad \tau Q_M^\tau = \delta Q_H^\delta. \end{aligned} \quad (4.10)$$

### Kaldor-Hicks Efficiency among Residents

To achieve Kaldor-Hicks efficiency among residents, the planner has decided to maximize the total welfare of the residents, albeit with the possibility of causing some utility loss for those in MR:

$$\begin{aligned} \max_{\tau} \left( \max_{\delta} \left( v_H^\delta + v_M^\tau \right) \right), \\ \text{s.t. } \tau Q_M^\tau = \delta Q_H^\delta. \end{aligned} \quad (4.11)$$

## 4.3 Insurance with Extreme Loss

Despite evidence suggesting spatial dependency of climate risks ([Tack and Ubilava, 2015](#); [Ker et al., 2015](#)), we assume the risk  $\epsilon_H$  and  $\epsilon_M$  are independent to simplify our model and focus on the core mechanics of the proposed redistribution scheme.

### 4.3.1 Risk

To simplify the representation of extreme loss, we have categorized  $\epsilon_H$  into three possible states,  $c_H \in \{0, l, L\}$ , and  $\epsilon_M$  into four states,  $c_M \in \{00, l0, 0L, lL\}$ . In HR, 0 represents no loss incurred,  $l$  a moderate loss, and  $L$  a large loss that triggers PDR. For MR, 00 denotes no loss in MR and no large loss in HR, with PDR not triggered,  $l0$  denotes a moderate loss in MR without a large loss in HR, also with PDR not triggered,  $0L$  denotes no loss in MR but a large loss in HR, triggering PDR, and  $lL$  denotes both a moderate loss in MR and a large loss in HR, triggering PDR. Four states are used to represent the situations in MR as we assume the moderate loss in MR is independent from large loss HR.

#### High Risk Region Risk

Within the three states framework, the risk of HR manifests itself in one of three distinct scenarios. The most severe outcome is an extreme loss, symbolized by  $L$ , which, if it occurs, results in a significant loss uniformly affecting all residents. The likelihood of this extreme event is represented by the probability  $r$ . This scenario underscores the potential for widespread impact under certain conditions, emphasizing the critical nature of planning for and mitigating such risks.

In situations where the extreme loss does not occur, the model shifts focus to more common but less severe outcomes. These include scenarios where residents either face no loss at all or incur a moderate loss  $l$ . The transition between no loss and this moderate state under normal circumstances is governed by the probability  $p$ . The moderate risk outcome may vary across different residents.

In the state of  $L$ , residents in HR will receive a wealth transfer financed in MR through the PDR schemes to cover the unprotected loss. Under this three-states framework, the actuarially fair premium of  $\epsilon_H$  is  $p(1 - r)l + rL$ . Define the premium loading factor  $\lambda_H = \pi_H / (p(1 - r)l + rL)$ . For  $\lambda_H > 1$ ,  $\lambda_H = 1$ , and  $\lambda_H < 1$ , the premium will be actuarially unfair, actuarially fair, and actuarially favorable, respectively.

In the case where a resident of HR experiences an extreme loss, a wealth transfer from

MR to HR will be triggered through the PDR scheme. The wealth of the three states, with and without the existence of insurance coverage, is displayed in Table 4.2 Panel A.

**TABLE 4.1**  
**Joint Probability Structure of HR Risk**

	No Moderate Loss	Moderate Loss	
Moderate Scenario	0	$l$	$1 - r$
Extreme Scenario	$L$		$r$
	$1 - p$	$p$	

### Moderate Risk Region Risk

We model the risk in MR using a two-point distribution, governed by a probability  $q$  of incurring a loss. In the states 00 and 0L, residents within the MR region experience no loss. Conversely, in the states  $l0$  and  $lL$ , they endure a moderate loss represented by  $l$ . Consequently, the actuarially fair premium for MR residents is  $ql$ . Similar to the HR scenario, we define  $\lambda_M = \pi_M/(ql)$  to evaluate pricing fairness.

Unlike HR, MR is not susceptible to extreme losses. However, MR faces a contingent liability for wealth transfer via PDR should an extreme loss event occur in HR. Specifically, when HR's residents encounter an extreme loss—occurring with a probability  $r$  as described in Section 4.3.1—MR's residents then bear part of the burden, manifesting in states 0L and  $lL$ . This liability takes the form of ex-post taxation, valued at a present value of  $\theta$ , designed to compensate for the financial shortfall from HR's catastrophic loss.

The wealth of MR in the four states is displayed in Table 4.2 Panel B.

## 4.3.2 Insurance Demand Function

### High Risk Region Residents' Demand Function

Given the premium  $\pi_H$ , the residents' objective is to choose a level of coverage  $Q_H \geq 0$  that maximizes their expected utility (4.6). A first-order condition for an interior solution

**TABLE 4.2**  
**Discrete States Framework**

<i>Panel A: High Risk Region</i>				
State $c_H$	0	$l$	$L$	
Probability	$(1-p)(1-r)$	$p(1-r)$	$r$	
Wealth, no insurance cover	$w$	$w-l$	$w-L$	
Wealth, insurance cover at $Q_H$	$w-\pi_H$	$w-\pi_H-l+Q_Hl$	$w-\pi_H-L+Q_HL+(1-Q_H)\theta$	
<i>Panel B: Moderate Risk Region</i>				
State $c_M$	00	l0	0L	lL
Probability	$(1-q)(1-r)$	$q(1-r)$	$(1-q)r$	$qr$
Wealth, no insurance cover	$w$	$w-l$	$w-(1-Q_H)\theta$	$w-l-(1-Q_H)\theta$
Wealth, insurance cover at $Q_M$	$w-\pi_M$	$w-\pi_M-l+Q_Ml$	$w-\pi_M-(1-Q_H)\theta$	$w-\pi_M-l+Q_Ml-(1-Q_H)\theta$

to this program is

$$\pi_H(1-r)(1-p)u'_0 + (\pi_H-l)(1-r)pu'_l + (\pi_H+\theta-L)ru'_L = 0. \quad (4.12)$$

Here, for the sake of notation convenience,  $u'_c$  denotes the marginal utility in state  $c$ . Expected utility is concave in  $Q_H$ , so the second order condition is trivially satisfied. The solution  $(Q_H, \pi_H)$  of the first-order condition (4.12) forms a demand function of the residents in HR, with  $\partial Q_H/\partial \pi_H < 0$ . The following proposition states the upper bound of residents' optimal coverage demand:

**Proposition 4.1.** *With  $\theta \geq 0$ , the optimal coverage  $Q_H \leq 1$  when the premium pricing is actuarially fair, and strictly less than 1 when  $\theta > 0$ .*

Contrary to the well-established result in insurance literature, which states that a risk-averse individual offered an actuarially fair premium will opt for full coverage, Proposition 4.1 suggests that the existence of PDR leads residents of HR areas to demand less than full coverage, even at an actuarially fair premium.

The following proposition characterizes the demand for insurance coverage for PDR scale:

**Proposition 4.2.** *For each level of premium ratio  $\pi_H$ , the optimal coverage ratio for HR*

residents  $Q_H$  decreases with the scale of public disaster relief  $\theta$ , and is strictly decreasing if  $Q_H > 0$ .

The rationale behind Proposition 4.1 and 4.2 is that PDR effectively mitigates the losses of HR residents during extreme events without requiring a premium, thereby altering their incentive to seek full insurance coverage. Proposition 4.1 and 4.2 explain the phenomenon wherein the charity hazard leads to a reduction in the demand for climate insurance.

Stemmed from Proposition 4.2 and Proposition 4.4, we have the following corollary.

**Corollary 4.1.**  $\exists \theta' > 0$  and  $L' > 0$ , such that when  $\theta \leq \theta'$  and  $L \geq L'$ , residents' demand is more elastic to premium in MR than in LR for coverage ratio  $0 < Q < 1$ , i.e.,  $|E_{DM}(Q)| > |E_{DH}(Q)|$ .

The intuition behind this corollary is that when the PDR offers coverage that only partially mitigates the extreme risks faced by residents in HR, their fundamental risk exposure remains largely unchanged. Consequently, the demand for insurance among these residents shows lower elasticity in response to increases in premiums due to their higher expected losses and greater volatility.

## HR Demand Function in the Case of Constant Absolute Risk Aversion

If the utility function satisfies constant absolute risk aversion, that is,  $u' \propto e^{-\gamma x}$  for some coefficient of absolute risk aversion  $\gamma$ , the first-order condition (4.12) can be rearranged to give the explicit demand function that maps the optimal coverage ratio to premium rates.

**Lemma 4.1.** For a resident in HR with constant absolute risk aversion of  $\gamma > 0$ , her demand function  $D_H : Q_H \rightarrow \pi_H$  is

$$D_H(Q_H; \gamma) = \frac{(1-r)pl \exp[(1-Q_H)l\gamma] + r(L-\theta) \exp[(1-Q_H)(L-\theta)\gamma]}{(1-r)(1-p) + (1-r)p \exp[(1-Q_H)l\gamma] + r \exp[(1-Q_H)(L-\theta)\gamma]}. \quad (4.13)$$

## Moderate Risk Region Residents' Demand Function

Similar to the HR residents' scenario, the MR residents' objective is to choose a level of coverage  $Q_M \geq 0$  that maximizes their expected utility (4.7). Given that HR residents'

coverage decision is not affected by the decisions of MR, the optimal coverage ratio  $Q_H^*$  is taken as exogenously given for the residents in MR. A first-order condition for an interior solution to this program is

$$\pi_M(1-r)(1-q)u'_{00} + (\pi_M - l)(1-r)qu'_{l0} + \pi_M r(1-q)u'_{0L} + (\pi_M - l)rq u'_{lL} = 0, \quad (4.14)$$

where  $u'_c$  denotes the marginal utility in state  $c$ . Expected utility is concave in  $Q_M$ , so the second order condition is trivially satisfied. The solution  $(Q_M, \pi_M)$  of the first-order condition (4.14) forms a demand function for the residents in MR, with  $\partial Q_M / \partial \pi_M < 0$ . The following proposition states the characteristics of residents' optimal coverage demand in terms of pricing fairness:

**Proposition 4.3.** *With  $\theta \geq 0$ , the optimal coverage ratio  $Q_M = 1$  when the premium pricing is actuarially fair.*

Contrary to residents in HR, insurance demand among MR residents does not escalate with an increase in the scale of PDR. The underlying rationale is that while an increase in  $\theta$  enhances the wealth contingency for MR residents, it does not address a risk that traditional insurance against localized risks can mitigate. Consequently, although an increase in  $\theta$  decreases the expected utility for MR residents, it does not stimulate an increased demand for insurance.

The following proposition characterises the demand for insurance coverage for post disaster relief scale, when premiums are actuarially unfair.

**Proposition 4.4.** *When the pricing is actuarially unfair, the optimal coverage ratio  $Q_M$  is increasing in the scale of public disaster relief  $\theta$  if two conditions are satisfied. (1) the residents are *risk prudent*, indicated by*

$$u'''(\cdot) > 0.$$

(2) the *premium loading factor* satisfies

$$\lambda_M \leq \frac{u''_{lL}}{(1-q)u''_{0L} + qu''_{lL}}. \quad (4.15)$$

Otherwise, the  $Q_M$  is strictly decreasing as  $\theta$  increases.

This proposition states that residents who are risk-prudent show a greater willingness to manage additional risks under favorable financial conditions, specifically in states 00 and 0L. Consequently, their demand for insurance is expected to increase as the PDR scale increases.

### MR Demand Function in the Case of Constant Absolute Risk Aversion

In the case where the residents' utility is constant absolute risk aversion, we can obtain an explicit form of the demand function.

**Lemma 4.2.** *For a resident in MR with constant absolute risk aversion of  $\gamma > 0$ , her demand function  $D_M : Q_M \rightarrow \pi_M$  is*

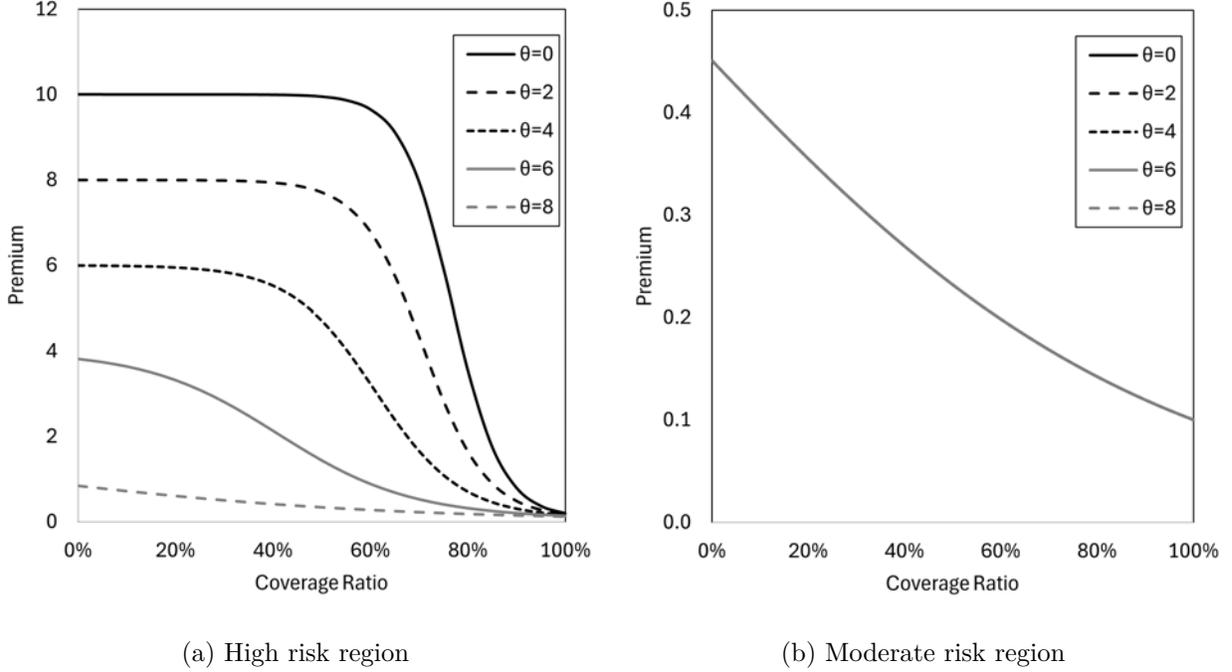
$$D_M(Q_M; \gamma) = \frac{ql \exp[(1 - Q_M)l\gamma]}{(1 - q) + q \exp[(1 - Q_M)l\gamma]}. \quad (4.16)$$

Notice that the demand function for MR is independent of  $\theta$ , which represents a special case of Proposition 4.4. When the resident of MR follows constant absolute risk aversion utility, both sides of condition (4.15) are equal. This can be easily verified by substituting  $\pi_M = \lambda_M ql$  into the left-hand side of Equation (4.16).

The HR and MR demand functions in the case of constant absolute risk aversion are depicted in Figure 4.3.

#### 4.3.3 Insurance Supply Function

In our discrete states framework, it is reasonable to assume that the probability of incurring a loss ( $\{l, L\}$  for HR and  $\{l0, lL\}$  for MR) exceeds the solvency requirement threshold  $\alpha$  (i.e.,  $1 - (1 - r)(1 - p) > \alpha$  and  $q > \alpha$ ). If this were not the case, the required solvency reserve for insurers would be zero. This scenario would allow insurers to profit from premiums in loss-free states, but in the event of any loss, they would face bankruptcy, transferring the remaining liabilities to the planner and thus defeating the purpose of private sector participation.



**Figure 4.3.** Coverage demand curve of both region under constant absolute risk aversion utility.

Our analysis focuses solely on the rational supply from domestic insurers operating in either region. This approach is taken because the complexities of the internal capital markets of multi-regional insurers would significantly complicate our model<sup>6</sup>.

Denote  $D = \frac{1}{1+r_f}$  as the non-contingent discounting factor. The following statement characterizes the supply curve of insurers operating in HR and MR under different risk scenarios:

**Proposition 4.5.** *For risk-neutral insurers who have the objective function (4.5), their pre-redistribution insurance supply in HR and MR is described as follows:*

- In MR, the supply remains perfectly elastic at the premium rate of  $\pi_M(\beta) = lq + lD(\beta - r_f)$  up to a coverage level of  $\bar{Q}_M(v) = v/(lD)$ , beyond which it becomes perfectly inelastic.

<sup>6</sup>A firm's internal capital market is a mechanism within a conglomerate or multi-divisional company that facilitates the redistribution of capital and resources among its various divisions or subsidiaries, based on their potential returns and strategic importance (Stein, 1997; Shin and Stulz, 1998). Research indicates that insurers operating both life and P&C divisions often increase premiums in the life division when the P&C division experiences unexpected catastrophic shocks (Ge, 2021). This strategic use of the firm's internal capital market helps stabilize the performance of affected divisions. Insurers with business across multiple regions could similarly manage premium requirements and insurance coverage supply decisions.

- In HR, if  $r \geq \alpha$ , the supply remains perfect elastic at the premium rate of  $\underline{\pi}_H(\beta) = l(1-r)p + Lr + LD(\beta - r_f)$  up to a coverage level of  $\overline{Q}_H(v) = v/(LD)$  beyond which it becomes perfectly inelastic.
- In HR, if  $r < \alpha$ , the supply remains perfect elastic at the premium rate of  $\underline{\pi}_H(\beta) = l(1-r)p + lr + Lr + (\beta - r_f)D[l + r(L-l)]$  up to a coverage level of  $\overline{Q}_H(v) = v/[D(rL + (1-r)l)]$  beyond which it becomes perfectly inelastic.

The supply curve for insurers in HR and MR features a critical turning point, denoted as  $(\overline{Q}_H(v), \underline{\pi}_H(\beta))$  for HR and  $(\overline{Q}_M(v), \underline{\pi}_M(\beta))$  for MR, respectively. As the risks are exogenously given, this turning point is determined by the characteristics of the insurer, including her initial capital  $v$  and her required rate of return  $\beta$ . The terms  $\overline{Q}_H(v)$  and  $\underline{\pi}_H(\beta)$  represent the maximum coverage an insurer with  $(\beta, v)$  can offer and the minimum premium required by her, respectively. Similarly,  $\overline{Q}_M(v)$  and  $\underline{\pi}_M(\beta)$  denote the corresponding maximum coverage and minimum premium for an insurer with  $(\beta, v)$  operating in MR. For the sake of notation simplicity, hereafter we denote  $(\overline{Q}_H(v), \underline{\pi}_H(\beta))$  and  $(\overline{Q}_M(v), \underline{\pi}_M(\beta))$  as  $(\overline{Q}_H, \underline{\pi}_H)$  and  $(\overline{Q}_M, \underline{\pi}_M)$ .

Proposition 4.5 illustrates that a public insolvency guarantee allows the planner to define the level of public sector involvement in mitigating climate risk by setting appropriate solvency requirements,  $\alpha$ .

When  $\alpha$  is set smaller than catastrophe probability, such that  $r \geq \alpha$ , the public sector carries limited catastrophe loss liability (or, in our discrete states model, no liability at all), placing the burden of HR's climate risk and the obligation to meet solvency capital requirements entirely on the private sector. The capital requirement for full coverage is therefore  $L$ , and the insurer will demand an additional premium on top of the net premium to meet the required rate of return, whose present value at  $t = 0$  is  $LD(\beta - r_f)$ . The maximum coverage that the insurer can provide is given by  $\overline{Q}_H(v) = v/(LD)$ .

Conversely, by choosing an  $\alpha$  greater than catastrophe probability, where  $r < \alpha$ , the public sector effectively acts as a provider of excess loss reinsurance to the private sector. Consequently, this reduces the minimum required premium  $\underline{\pi}_H(\beta)$  and increases the maximum coverage  $\overline{Q}_H(v)$ .

Following Proposition 4.5, we have the following corollary.

**Corollary 4.2.** *For  $0 \leq \bar{v}_0 < \infty$ , insurers' supply is more elastic to premium in MR than in LR for coverage ratio  $0 < Q < 1$ , i.e.,  $|E_{SM}(Q)| > |E_{SH}(Q)|$ .*

This corollary says that the supply function is more elastic in the MR than in HR within the feasible range of coverage ratio.

Building upon the supply functions outlined in Proposition 4.5, the subsequent theorem describes the availability of insurance coverage in both HR and MR.

**Theorem 4.1.** *For an insurer characterized by a required rate of return  $\beta \geq r_f$  and an initial endowment  $v$ , and when  $p = q$ , it is established that  $\bar{Q}_H < \bar{Q}_M$  and  $\underline{\pi}_H > \underline{\pi}_M$ , irrespective of whether  $r \geq \alpha$  or  $r < \alpha$ .*

This theorem illustrates that when resources are identical, the presence of climate risk results in a less favorable insurance supply for residents in HR compared to those in MR regions. This disparity manifests as a more unfavorable lower bound for premium requirements and reduced coverage availability in HR areas, both consequences of the increased challenge climate risk poses to meeting solvency capital requirements.

For traceability purposes, we assume insurers are risk-neutral in this chapter. However, Theorem 4.1 still holds when insurers are risk-averse. Unlike the transition from perfect elasticity to perfect inelasticity at the turning point, the supply curve instead exhibits a gradual increase with diminishing rates of growth on both sides of the turning point.

To utilize the fact that the supply curve is simpler than the demand curve, in this chapter, we focus on tax redistribution from the insurer's side.

#### 4.3.4 Optimal Redistribution

Recall that we propose using the collected tax to subsidize premiums on policies written and to increase capacity in HR. For a tax redistribution scheme denoted as  $(\tau, \delta)$ , the distribution of collected taxes is categorized into two methods:  $\psi \geq 0$  for premium subsidy and  $\phi \geq 0$  for providing additional capacity, such that  $\delta = (\psi, \phi)$ .

Before discussing the optimal form of tax redistribution, we first establish the relationship between these two uses of the collected tax and their impact on insurance supply in HR:

- For per coverage provided by insurer, denote  $\phi$  as the additional coinsurance provided by the planner. The turning point of supply curve is increased to  $\bar{Q}_H^\delta = (1 + \phi)\bar{Q}_H$ .
- For per coverage provided by insurer, denote  $\psi$  as the premium subsidy rate provided by the planner. The required premium before the turning point is reduced to  $\underline{\pi}_H^\delta = \underline{\pi}_H - \psi$ .

Recall that the tax rate per coverage unit provided in MR is denoted as  $\tau$ , and the post-taxation coverage equilibrium in MR is  $Q_M^\tau$ . In cases where the subsidy and coinsurance are fully financed by the tax revenues from coverage provided in MR, the total tax expenditure in HR is limited to the tax collection amounting to  $\tau Q_M^\tau$ .

Given the tax collected, the optimal tax redistribution plan is outlined in the following theorem:

**Theorem 4.2.** *The post-redistribution supply curve  $S_H^{\delta^*}(Q)$  of the optimal redistribution plan  $\delta^* = (\psi^*, \phi^*)$  intersects the demand curve of HR residents  $D_H(Q)$  at the turning point of  $S_H^{\delta^*}(Q)$ , i.e.  $(\bar{Q}_H^{\delta^*}, \underline{\pi}_H^{\delta^*})$ .*

*The optimize redistribution plan that maximizes the expected utility of residents in HR, denoted as  $\delta^* = (\psi^*, \phi^*)$ , solves the following equations*

$$\begin{cases} \phi &= \frac{\tau Q_M^\tau - \psi(1 + \phi)\bar{Q}_H}{rL + (1 - r)pl}, \\ \psi &= \underline{\pi}_H - D_H((1 + \phi)\bar{Q}_H). \end{cases} \quad (4.17)$$

The optimal redistribution solution presented in Theorem 4.2 is expressed implicitly. Obtaining a closed-form solution would require detailed knowledge of the demand function for residents in the high-risk region, which depends on their utility functions. We will not pursue this further because making explicit assumptions about these utility functions could introduce biases into the theorem's conclusions.

## 4.4 Conclusion Remarks

In this chapter, we expand on the concept of a private-public partnership by introducing a self-financing tax redistribution framework. This framework levies taxes on the excess insurance coverage in regions of moderate risk and reallocates these funds to high-risk regions in the form of co-insurance provision and premium subsidies. While this redistribution scheme directly benefits residents in high climate risk regions, it also potentially increases overall utility for those in moderate-risk regions by mitigating climate risk externalities. Utilizing public disaster relief as an example of climate externalities, which is effectively a wealth transfer from low- to high-risk regions, we develop a discrete states model to capture the essence of climate risk. Through this model, we illustrate the challenges faced by domestic insurers operating in high-risk areas and propose an optimal redistribution plan based on predetermined tax collection to help improve their situation. The analyses underscore the importance of accounting for climate externalities in risk-sharing arrangements and modeling willingness to pay, offering a novel approach to closing the climate risk protection gap within a PPP framework.

## 4A Appendix

### 4A.1 Proofs

**Proof for Proposition 4.1** By substituting  $Q_H = 1$  into the implicit demand function (4.12), we conclude the condition for full coverage is  $\pi_H \leq pl(1-r) + (L-\theta)r$ , which is less than or equal to the actuarially fair premium of  $pl(1-r) + Lr$  when  $\theta \geq 0$ , resulting optimal coverage ratio demand from HR residents is always less than or equal to 1 when the premium pricing is actuarially fair.

Q.E.D.

**Proof for Proposition 4.2** For the sake of simplifying notation, we denote the first-order condition (4.12) as  $f(Q_H, \theta) = 0$ . By implicit function theorem, the demand for insurance coverage for PDR scale can be obtained from

$$\frac{\partial Q_H}{\partial \theta} = -\frac{\frac{\partial f}{\partial \theta}}{\frac{\partial f}{\partial Q_H}}, \quad (4A.1)$$

where

$$\frac{\partial f}{\partial Q_H} = \pi_H^2(1-r)(1-p)u_0'' + (\pi_H - l)^2(1-r)pu_l'' + (\pi_H + \theta - L)^2ru_L'', \quad (4A.2)$$

and

$$\frac{\partial f}{\partial \theta} = ru_L' - (L - \theta - \pi_H)ru_L'' \quad (4A.3)$$

are both strictly negative, given that  $u'(\cdot) > 0$  and  $u''(\cdot) < 0$ . Thus,  $\frac{\partial Q_H}{\partial \theta} < 0$ . As  $Q_H$  is lower bounded at zero, when  $Q_H = 0$ , the demand for insurance no longer decreases as  $\theta$  increases.

Q.E.D.

**Proof for Proposition 4.3** When  $Q_M = 1$ , the implicit demand function (4.14) become

$$\begin{aligned} & \pi_M(1-r)(1-q)u'(w-\pi_M) + (\pi_M-l)(1-r)qu'(w-\pi_M) + \\ & \pi_M r(1-q)u'(w-\pi_M - (1-Q_H)\theta) + (\pi_M-l)rqu'(w-\pi_M - (1-Q_H)\theta) = 0. \end{aligned}$$

With some derivations, we have  $\pi_M = lq$ . The coverage demand premium ratio is equal to the actuarially fair premium ratio of  $ql$  when  $\theta \geq 0$ , resulting optimal coverage ratio demand from HR residents is equal to 1 when the premium pricing is actuarially fair.

Q.E.D.

#### Proof for Proposition 4.4

**Proof** For the sick of simplifying notation, we denote the first-order condition (4.14) as  $f(Q_M, \theta) = 0$ . By implicit function theorem, the demand for insurance coverage for PDR scale can be obtained from

$$\frac{\partial Q_M}{\partial \theta} = -\frac{\frac{\partial f}{\partial \theta}}{\frac{\partial f}{\partial Q_M}}, \quad (4A.4)$$

where

$$\frac{\partial f}{\partial Q_M} = \pi_M^2(1-r)(1-q)u''_{00} + (\pi_M-l)^2(1-r)qu''_{l0} + \pi_M^2 r(1-q)u''_{0L} + (\pi_M-l)^2 rqu''_{lL}, \quad (4A.5)$$

is strictly negative, given that  $u''(\cdot) < 0$ . Therefore, the relation of  $Q_M$  and  $\theta$  is depend on the sign of

$$\frac{\partial f}{\partial \theta} = \pi_M r(1-q)u''_{0L} + (\pi_M-l)rqu''_{lL}. \quad (4A.6)$$

When  $\partial f/\partial \theta$  is positive,  $Q_M$  strictly increases as  $\theta$  increases, which requests

$$\frac{\pi_M(1-q)}{(l-\pi_M)q} < \frac{u''_{lL}}{u''_{0L}}. \quad (4A.7)$$

Substituting  $\pi_M = \lambda_M ql$ , we have

$$\frac{\lambda_M(1-q)}{(1-\lambda_M q)} < \frac{u''_{lL}}{u''_{0L}}. \quad (4A.8)$$

Given that the pricing is actuarially unfair, which means  $\lambda_M > 1$ , it follows that the left-hand side of inequality (4A.8) exceeds 1. For this inequality to be valid, it is necessary for residents to exhibit risk-prudent behavior, characterized by  $u''(\cdot) < 0$  and  $u'''(\cdot) > 0$ . Otherwise, the right-hand side of inequality (4A.8) would be strictly less than 1, rendering the condition unsatisfied.

Notably, the condition  $(1 - \lambda_M q) > 0$  is trivially satisfied; if not,  $\pi_M$  would be greater than or equal to  $l$ , which would undermine the rationale for demanding insurance. After further derivations, we obtain the second condition as detailed in Equation (4.15).

Q.E.D.

**Proof for Proposition 4.5** For a risk-neutral insurer operating in a MR region, given that  $q > \alpha$ , her VaR at the  $\alpha$ -level is  $l$ . Within the framework of the discrete states model, she is required to reserve solvency capital amounting to  $v_0 = lQ_M/(1 + r_f)$  to underwrite this coverage. With such capital reservation, her risk of insolvency is nullified, resulting in  $\pi_i = 0$ . The insurer's objective function, as detailed in equation (4.5) under the discrete states model, is presented as follows

$$\begin{aligned} \max_{Q_M} \quad & (1 + \beta) \left( v - \frac{lQ_M}{1 + r_f} \right) + \pi_M Q_M - lqQ_M + lQ_M, \\ \text{s.t.} \quad & 0 \leq \frac{lQ_M}{1 + r_f} \leq v. \end{aligned} \tag{4A.9}$$

The first-order condition is

$$-(1 + \beta) \frac{l}{1 + r_f} + \pi_M - lq + l = 0. \tag{4A.10}$$

After deriving and substituting  $D = 1/(1 + r_f)$ , we arrive at the formula  $\underline{\pi}_M(\beta) = lq + lD(\beta - r_f)$ . The maximum coverage an insurer can offer occurs when the solvency capital requirement equals her initial endowment, which is achieved at  $\bar{Q}_M = v/(lD)$ .

For a risk-neutral insurer operating in a HR region, when  $1 - (1 - r)(1 - p) \geq \alpha$  and  $r \geq \alpha$ , her VaR at the  $\alpha$ -level is  $L$ . Within the framework of the discrete states model,

she is required to reserve solvency capital amounting to  $v_0 = LQ_H/(1 + r_f)$  to underwrite this coverage. With such capital reservation, her risk of insolvency is nullified, resulting in  $\pi_i = 0$ . The insurer's objective function, as detailed in equation (4.5) under the discrete states model, is presented as follows

$$\begin{aligned} \max_{Q_H} \quad & (1 + \beta) \left( v - \frac{LQ_H}{1 + r_f} \right) + \pi_H Q_H - lp(1 - r)Q_H - LrQ_H + LQ_H, \\ \text{s.t.} \quad & 0 \leq \frac{LQ_H}{1 + r_f} \leq v. \end{aligned} \quad (4A.11)$$

The first-order condition is

$$-(1 + \beta) \frac{L}{1 + r_f} + \pi_H - lp(1 - r) - Lr + L = 0. \quad (4A.12)$$

After deriving and substituting  $D = 1/(1 + r_f)$ , we arrive at the formula  $\pi_H(\beta) = l(1 - r)p + Lr + LD(\beta - r_f)$ . The maximum coverage an insurer can offer occurs when the solvency capital requirement equals her initial endowment, which is achieved at  $\bar{Q}_H = v/(LD)$ .

For a risk-neutral insurer operating in a HR region, when  $1 - (1 - r)(1 - p) \geq \alpha$  and  $r < \alpha$ , her VaR at the  $\alpha$ -level is  $l$ . Within the framework of the discrete states model, she is required to reserve solvency capital amounting to  $v_0 = lQ_H/(1 + r_f)$  to underwrite this coverage. With such capital reservation, her insolvency risk premium based on equation (4.4) is  $\pi_i = r(L - l)Q_H/(1 + r_f)$ . The insurer's objective function, as detailed in equation (4.5) under the discrete states model, is presented as follows

$$\begin{aligned} \max_{Q_H} \quad & (1 + \beta) \left( v - \frac{lQ_H}{1 + r_f} - \frac{r(L - l)Q_H}{1 + r_f} \right) + \pi_H Q_H - lp(1 - r)Q_H - lrQ_H + l(1 - r)Q_H, \\ \text{s.t.} \quad & 0 \leq \frac{lQ_H}{1 + r_f} + \frac{r(L - l)Q_H}{1 + r_f} \leq v. \end{aligned} \quad (4A.13)$$

The first-order condition is

$$-(1 + \beta) \frac{l}{1 + r_f} - (1 + \beta) \frac{r(L - l)}{1 + r_f} + \pi_H - lp(1 - r) - lr + l(1 - r) = 0. \quad (4A.14)$$

After deriving and substituting  $D = 1/(1 + r_f)$ , we arrive at the formula  $\pi_H(\beta) = l(1 - r)p + lr + Lr + (\beta - r_f)D[l + r(L - l)]$ . The maximum coverage an insurer can offer occurs when the solvency capital requirement equals her initial endowment, which is achieved at  $\bar{Q}_H(v) = v/[D(rL + (1 - r)l)]$ .

Q.E.D.

**Proof for Theorem 4.1** When  $r \geq \alpha$ , it follows naturally from  $L > l$  that  $v/(LD) < v/(lD)$ , leading to the conclusion that  $\bar{Q}_H < \bar{Q}_M$ . To demonstrate that  $\pi_H > \pi_M$ , we need to establish the following inequality:

$$l(1 - r)p + Lr + LD(\beta - r_f) > lq + lD(\beta - r_f).$$

Upon derivation, this simplifies to:

$$r > \frac{l(q - p) - (L - l)D(\beta - r_f)}{L + l},$$

which is satisfied under the conditions  $p = q$ ,  $\beta \geq r_f$ , and  $r \geq 0$ .

When  $r < \alpha$ , it follows naturally from  $L > l$  that  $v/[D(rL + (1 - r)l)] < v/(lD)$ , leading to the conclusion that  $\bar{Q}_H < \bar{Q}_M$ . To demonstrate that  $\pi_H > \pi_M$ , we need to establish the following inequality:

$$l(1 - r)p + lr + Lr + (\beta - r_f)D[l + r(L - l)] > lq + lD(\beta - r_f).$$

Upon derivation, this simplifies to:

$$r > \frac{l(q - p) - (L - l)D(\beta - r_f)}{l(1 - p) + L + (L - l)D(\beta - r_f)},$$

which is satisfied under the conditions  $p = q$ ,  $\beta \geq r_f$ , and  $r \geq 0$ .

Q.E.D.

**Proof for Theroem 4.2** Given that the planner is risk-neutral and not subject to insolvency risk, the tax allocated to finance the premium for additional co-insurance is calculated as  $\phi$  times the fair premium for the HR region's background risk, denoted by  $\phi(rL + (1-r)pl)$ . Concurrently, the tax allocated for subsidizing coverage is expressed as  $\psi Q_H^\delta = \psi(1 + \phi)\bar{Q}_H$ . The aggregate of these two components is limited by the total tax revenue collected from the MR region:

$$\tau Q_M^\tau = \phi(rL + (1-r)pl) + \psi(1 + \phi)\bar{Q}_H.$$

Deriving from this, we establish the first equation that guides the implicit optimal redistribution strategy (4.17).

In this framework, the equilibrium following redistribution is identified at the juncture where the HR demand curve,  $D_H(Q)$ , intersects with the post-redistribution supply curve,  $S_H^{\delta^*}(Q)$ . As a result, the adjusted premium post-redistribution is captured by  $\underline{\pi}_H^{\delta^*} = D_H((1 + \phi)\bar{Q}_H)$ , and the subsidy ratio is determined by the difference between  $\underline{\pi}_H$  and  $\underline{\pi}_H^{\delta^*}$ :

$$\psi = \underline{\pi}_H - D_H((1 + \phi)\bar{Q}_H),$$

yielding the second equation for the implicit optimal redistribution strategy (4.17).

Q.E.D.

# Chapter 5

## Resilience Through Responsibility: The Impact of ESG Status on Firm Resilience to Climate-Induced Disruptions

### 5.1 Introduction

Global annual economic losses from Natural Catastrophes (NatCats) have shown a near-continuous increase over recent decades. Studies show that annual economic losses from NatCats exceeded \$350 billion in both 2011 and 2017, maintaining a consistent baseline of over \$100 billion annually. Before 2010, such losses had never surpassed the \$300 billion threshold.<sup>1</sup>

The insurance industry is likely the industry most affected by NatCats within the finance sector, as such events impact both the assets and liabilities of insurers. Consider life insurance companies for an example. It is well established that climate change can influence human life expectancy through both direct and indirect effects on health and economic development

---

<sup>1</sup>The statistics in use are retrieved from Munich Re's climate change study: <https://www.munichre.com/en/risks/climate-change.html>.

(See [Rocque et al., 2021](#), for a detailed review). This complexity makes premium pricing increasingly challenging. Additionally, life insurers are required to maintain substantial reserves with long durations to cover future claim obligations. The challenges posed by climate change on asset valuation further complicate the selection of assets for reserves, adding another layer of challenge for insurers. Beyond insurers, the impact of climate change on asset valuation is also a great concern for policymakers and institutional investors globally.

Recent studies have provided evidence that assets with exceptional Environmental, Social, and Governance (ESG) performance exhibit lower climate beta (i.e., the sensitivity of asset's value to changes in climate-related factors), suggesting a general trend of resilience to climate-related risks (e.g., [Pástor et al., 2021](#); [Bolton and Kacperczyk, 2021](#); [Krueger et al., 2020](#); [Choi et al., 2020](#); [Engle et al., 2020](#); [Hong et al., 2019](#); [Dietz et al., 2018](#)). This body of research underscores the potential of green assets (i.e., equities or credits issued by firms with good ESG practices) as a category that may offer relative stability in the face of climate change challenges.

The existing body of literature generally agrees that firms' ESG devotion is beneficial. From a risk management perspective, studies have highlighted how ESG initiatives contribute to a reduction in systematic risk (e.g., [El Ghouli et al., 2016](#); [Albuquerque et al., 2019](#)), [credit risk](#) (e.g., [Jiraporn et al., 2014](#)), and [downside risk](#) (e.g., [Hoepner et al., 2023](#); [Ilhan et al., 2019](#)). In terms of financing, firms with robust ESG profiles benefit from a lower [cost of capital](#) (e.g., [El Ghouli et al., 2011](#); [Chava, 2014](#); [Goss and Roberts, 2011](#)).

However, the discussion on how ESG contributions are reflected in share prices is ongoing, especially on the timing of value realization. Some researchers posit that ESG's value is quickly reflected at pivotal moments (e.g., [Deng et al., 2013](#); [Flammer, 2015](#); [Gillan, 2010](#)), suggesting an immediate market reaction to ESG advancements. Conversely, others argue for a gradual incorporation of ESG value into stock prices, unfolding over an extended period (e.g., [Edmans, 2011](#); [Dimson et al., 2015](#); [Lins et al., 2017](#); [Barko et al., 2017](#)), indicating a longer-term investment perspective on ESG's financial impacts. As the connection between ESG practices and firms' valuation is not straightforward, further exploration is needed to understand how ESG affects a company's financial health.

In this chapter, we conduct a comprehensive study of how ESG commitments impact a company's ability to withstand climate-related challenges. We test three potential explanations for why companies focused on sustainability tend to have lower climate risk: the market sentiment view, the representation of financial stability view, and the capacity for climate risk hedging view. After settling causality, our analysis establishes and reveals that the channel of investors' preference for green assets (market sentiment channel) offers the strongest explanation for the resilience of green assets against climate change. Moreover, in sectors most affected by climate change, the strategy of using investments to offset climate risks (climate risk hedging channel) also plays a significant role. However, the aspect of financial stability as a sign of resilience (financial stability proxy channel) is less significant in explaining why green assets are more resilient to climate change. This chapter contributes to the ongoing discussion about the value of ESG activities by illustrating how high ESG ratings enhance a firm's ability to recover from natural disasters, using these events as a natural experiment to demonstrate the tangible benefits of ESG commitments in enhancing resilience to external shocks.

A hurricane is a type of tropical cyclone, a powerful storm system with strong winds, heavy rainfall, and thunderstorms that forms over warm ocean waters. These storms can cause significant damage through high winds, storm surges, and flooding (Pielke Jr et al., 2008). Another common type of tropical cyclone is a typhoon. The main difference between a hurricane and a typhoon is their location: hurricanes occur in the Atlantic and Eastern Pacific Oceans, while typhoons occur in the Western Pacific Ocean. Climate change is expected to make hurricanes more intense (Balaguru et al., 2023) and more frequent (Lin et al., 2016), with stronger winds and increased rainfall due to rising sea surface temperatures. Additionally, higher sea levels may lead to more severe storm surges and coastal flooding, amplifying the overall impact of hurricanes (Strauss et al., 2021).

We categorize publicly traded firms on U.S. stock exchanges into three groups based on their exposure to hurricanes with damages exceeding 5 billion 2019 U.S. dollars: hurricane-impacted, hurricane-adjacent, and unimpacted firms. Utilizing Difference-in-Differences (DID) analysis, we uncover a significant relationship between firms' ESG practices and their

resilience to unexpected hurricane-related shocks, as evidenced by their Cumulative Abnormal Returns (CAR) during the period of hurricane impact. Further, to establish causal link, we conduct Regression Discontinuity Design (RDD) analysis, which leveraged the outcomes of shareholder proposal votes that narrowly passed or failed as an instrumental shock, examining its effect on firms' CAR within the hurricane impact window.

This chapter makes contribution to the climate finance literature. Prevailing research in climate asset pricing within the realm of climate finance has traditionally emphasized the long-term impacts of climate change, as seen in studies by [Addoum et al. \(2020\)](#), and [Baldauf et al. \(2020\)](#). Recent works have introduced another stream of literature exploring the short-term effects on market efficiency. In particular, [Schlenker and Taylor \(2021\)](#) uncover evidence that weather derivatives now possess the capability to mirror weather shocks for duration extending up to two weeks. Similarly, [Alok et al. \(2020\)](#) investigate whether fund managers exhibit overreactions to climate risks proximal to their firms, influenced by the salience and information effects. Their findings reveal that in the aftermath of a natural disaster occurring near a fund company, managers are likely to divest from companies based in the affected area at prices that undervalue these firms. This chapter contribute to the literature by reviewing that assets segregated by their ESG status exhibit varying degrees of climate resilience. It also reviews the mechanisms behind such segregation.

The remainder of this chapter is structured as follows: Section [5.2](#) discusses the three channels tested in this chapter and their rationales. Section [5.3](#) outlines the empirical methodology and data employed in this study. Section [5.4](#) reports our findings. Section [5.5](#) tests the robustness of our results. Section [5.6](#) concludes.

## 5.2 Potential Channels of Resilience Improvements

What are the potential reasons that firms with good ESG practices are more climate resilient towards catastrophe risk? In the following, we discuss three potential channels. (1) Market sentiment: Some investors derive [non-pecuniary utility](#) from investing ESG firms, and the presence of these ESG investors can reduce the likelihood of fire sales during hurricane

events. (2) Financial stability proxy: A firm’s ESG status may serve as a proxy for the firm’s financial health, suggesting that firms with higher ESG ratings could be less prone to financial distress in times of crisis. (3) Climate risk hedging: A firm’s ESG status can also indicate its climate hedging status. Firms with higher climate awareness may engage in more climate hedging activities, thereby improving their resilience to climate-related catastrophes.

### 5.2.1 Market Sentiment

Some investors prefer to invest in ESG firms. They often hold onto these investments, even during troubling events. [Pedersen et al. \(2021\)](#) and [Pástor et al. \(2021\)](#) point out that the ESG investors enjoy a non-pecuniary utility through holding the stocks issued by firms with good ESG status which alters the efficient frontier ([Pedersen et al., 2021](#)) and the equilibrium market return ([Pástor et al., 2021](#)).

Evidence suggests that brown assets have higher climate beta than green assets. [Engle et al. \(2020\)](#) find that using ESG scores to measure a company’s climate risk helps in creating investment portfolios that perform better in returns. [Choi et al. \(2020\)](#) observe that individual investors tend to sell off investments in companies with high carbon footprints during unusually hot weather. [Bolton and Kacperczyk \(2021\)](#) discover that investors expect higher returns from companies with greater carbon risks due to their potential negative impact on the environment. [Bansal et al. \(2016\)](#) point out that climate change is a significant risk for long-term investments. Finally, [Krueger et al. \(2020\)](#) note that large investors see climate risk as an important factor in making investment decisions.

The preference for ESG firms, coupled with their reduced climate beta, typically positions these assets as the last to be liquidated during climate crises. Additionally, the market sentiment channel suggests that investors generally favor industries with an ESG focus. However, this preference may lead to some ambiguity in the assessment of specific assets.

Should the resilience of green assets against natural catastrophes be attributable to behavioral and market sentiment, this hypothesis could be verified by examining the variation in holdings of green versus brown assets before and after NatCat events.

### 5.2.2 Financial Stability Proxy

ESG metrics may serve as indicators of a firm's financial health. A higher ESG score is associated with reduced systematic risk (El Ghouli et al., 2011; Albuquerque et al., 2019), lower downside risks (Hoepner et al., 2023; Ilhan et al., 2019), and diminished credit risks (Jiraporn et al., 2014; Seltzer et al., 2020). Furthermore, evidence suggests that firms with strong environmental credentials can secure financing at a more favorable cost of capital (El Ghouli et al., 2011; Chava, 2014; Goss and Roberts, 2011). Additionally, Luo and Balvers (2017) identified a premium associated with boycott risk.

In this chapter, we examine the financial stability proxy view by investigating whether a strong ESG status enhances a firm's long-term fundamentals following hurricane events. Alternatively, if financial distress is a significant factor enhancing the resilience of green assets during climate-related events, this hypothesis can be tested through a two-stage least square regression analysis. In the first step, the firm's ESG score is employed as the independent variable. Subsequently, financial distress, measured by Tobin's Q, is used as the instrumental variable. This approach allows for a detailed examination of the relationship between a firm's commitment to ESG principles and its resilience to financial distress in the context of climate events. We will schedule this analysis in future works.

### 5.2.3 Climate Risk Hedging

ESG metrics may serve as indicators of a firm's climate risk hedging strategies and concerns regarding climate risks. Zhang et al. (2023) have shown that firms with high ESG scores exhibit a greater managerial focus on long-term development, including an increased emphasis on resilience in the face of climate change. Companies more aware of climate risks are likely to engage in hedging activities, such as diversifying production, suppliers, and sales across geographically disparate regions to mitigate these risks.

To validate whether climate risk hedging effectively enhances resilience to disruptive climate events, one could examine the comparative climate resilience of firms that are sufficiently diversified against those that are not. This investigation would involve assessing

the impact of climate-related disruptions on both sets of firms to identify any significant differences in their resilience.

## 5.3 Methodology

### 5.3.1 Data

Our data are obtained from several sources. Stock data are from the daily Center for Research in Security Prices (CRSP) database, accounting data are from the Compustat Annually and Quarterly database, shareholders' proposal voting outcomes are from the RiskMetrics database (RiskMetrics), Hurricane event and loss data are from the Spatial Hazard Events and Losses Database v20.0 (SHELDUS), and ESG rating data are from MSCI ESG KLD database (KLD).

We use data of all U.S. common stocks traded on NYSE (New York Stock Exchange), AMEX (American Stock Exchange), and Nasdaq from 2003 to 2020 to construct stock-level CARs. This period was selected because prior to 2003, KLD only covered a small fraction of publicly traded firms (see Figure 5.1). We filter our sample by requiring all observations to have non-negative book equity, price equal to or greater than \$1, and at least three daily stock returns within a CAR window.

### 5.3.2 Key Variables Definition

#### Measure of a Firm's ESG Status

The KLD database encompasses approximately 650 companies before 2003 and expands to around 2,500 companies post-2003 (see Figure 5.1). It provides extensive data on firm-level ESG ratings across dimensions that are broadly categorized into Environmental, Social, and Governance sectors.

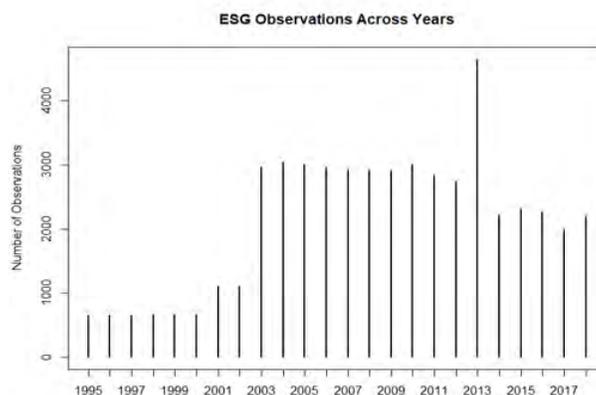
Following the methodology outlined by [Chen et al. \(2020\)](#), we selected 53 dimensions to assess a firm's ESG status. These dimensions cover areas such as environmental impact,

community engagement, workforce diversity, employee relations, product quality, and corporate governance. We evaluate each firm’s ESG performance by summing the scores in the “strengths” dimensions and subtracting the scores in the “concerns” dimensions.

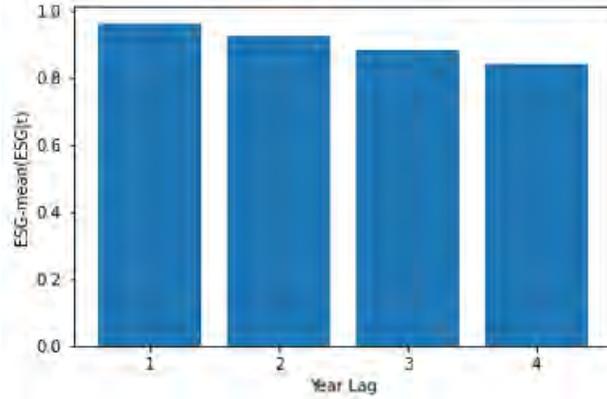
We observe a strong time-series correlation in firms’ ESG scores (see Figure 5.2). To avoid forward-looking bias in our analysis, we lag the firms’ ESG scores by one year in the subsequent regression analysis.

To derive separate scores for Environmental (E), Social (S), and Governance (G) components, we calculate the net scores specifically for each category. For the E score, we focus on environmental impact by subtracting the sum of “concerns” from the sum of “strengths” in this dimension. Similarly, for the S score, we aggregate the dimensions related to communities, workforce diversity, and employee relations. The G score is derived from evaluating product quality and corporate governance, ensuring a comprehensive assessment of a firm’s governance performance.

Table 5.1 presents the correlations among the composite ESG score and the individual E, S, and G scores. The composite ESG score shows the highest correlation with the S score (0.89), followed by the E score (0.58), and then the G score (0.32). Additionally, the E score is more correlated with the S score (0.25) than with the G score (0.10). The correlation between the S and G scores is notably low, at just 0.02.



**Figure 5.1.** Number of firms covered by MSCI ESG database by year.



**Figure 5.2.** Lagged correlation of ESG scores between year  $t$  and year  $t + k$  ( $k = 1$  to  $4$ ).

**TABLE 5.1**  
**Correlation Matrix of ESG, E, S, and G Scores**

This table presents the correlation matrix for Environmental (E), Social (S), Governance (G), and overall ESG scores. Each cell in the matrix displays the correlation coefficient between the respective variables, capturing the strength and direction of their linear relationship. Below each correlation coefficient, the  $p$ -values are provided in parentheses.

	E	S	G
ESG	0.5758 (0.0000)	0.8942 (0.0000)	0.3153 (0.0000)
E		0.2546 (0.0000)	0.1031 (0.0000)
S			0.0154 (0.0003)

## Measure of a Firm’s Resilience

We utilize an event study methodology to assess the impact of hurricane events on stock prices. In our study, we employ the cumulative abnormal return (CAR) to evaluate the resilience of firms’ valuations in response to hurricane events. Further, in Section 5.4.1, we conduct a univariate DID analysis using the CAR calculated for the period from the formation to the dissipation of the hurricane. We then compare these results with CARs derived from a control period, chosen arbitrarily, to understand the differential impact of hurricanes on firm valuation. The CAR, calculated over a specified window of days, is a widely used metric for assessing the short-term impact of an event on a firm’s valuation. Our calculation is detailed in Appendix 5A.2.

## Impacted Region, Adjacent Region, and Non-impact Region

Hurricanes pose a significant risk across a broad swath of the US, especially in coastal areas stretching from Texas to Maine. Not only do these regions face direct threats, but inland areas can also suffer from related flooding and strong winds.

The SHELDUS records 15 post-1980 hurricanes that made landfall in the continental U.S. that resulted in economic losses exceeding \$5 billion (in 2019 dollars). Detailed information about these hurricanes, including names, categories, and dates, alongside their economic impacts, is presented in Table 5.2. The table offers detailed information on each hurricane, including its name (Name), the Saffir-Simpson category (Category) it falls under, the year it occurred (Year), the date of formation (Start Date), the date it dissipated (End Date), the date it first made landfall (Landfall Date), the number of fatalities (Fatalities), and the cost of property damage adjusted to the 2019 dollar value (Damages, \$b). Focusing on hurricanes since 2003, we treat 10 significant hurricane events as pivotal in our analysis.

We categorize US counties based on their experience with each hurricane into three groups: impacted, adjacent, and non-impacted. Impacted counties are those with recorded losses for a specific hurricane in the SHELDUS database. Following Dessaint and Matray (2017), each affected county is matched with the five nearest counties, based on geographic

**TABLE 5.2****Hurricanes Utilized for Analyses: Characteristics and Impact**

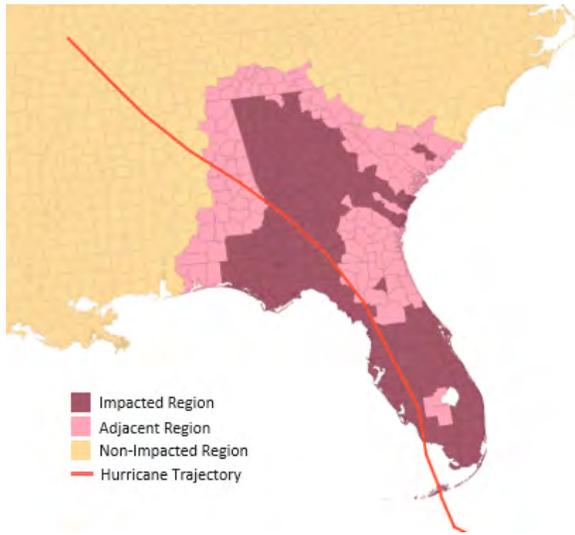
This table summarizes the hurricanes incorporated into the analyses as shocks, detailing key attributes and consequences of each event. For each hurricane, the table lists its name, Saffir-Simpson category, year of occurrence, event start and end dates, landfall date, fatalities, and damages. The damages are adjusted to reflect 2019 dollar values to ensure consistency in economic impact assessment across different years.

Name	Category	Year	Start Date	End Date	Landfall Date	Fatilities	Damages, \$b
Hugo	4	1989	10/9/1989	22/9/1989	22/9/1989	21	12.3
Andrew	5	1992	16/8/1992	28/8/1992	24/8/1992	26	41.2
Opal	3	1995	27/9/1995	5/10/1995	4/10/1995	9	7.4
Fran	3	1996	23/8/1996	8/9/1996	6/9/1996	26	5.8
Floyd	2	1999	7/9/1999	17/9/1999	14/9/1999	56	9
Alison	TS	2001	5/6/2001	17/6/2001	5/6/2001	41	11.1
Isabel	2	2003	6/9/2003	19/9/2003	18/9/2003	16	6.4
Charley	4	2004	9/8/2004	14/8/2004	13/8/2004	10	17.4
Frances	2	2004	25/8/2004	8/9/2004	5/9/2004	7	11
Ivan	3	2004	2/9/2004	24/9/2004	16/9/2004	25	21.7
Jeanne	3	2004	13/9/2004	28/9/2004	26/9/2004	4	8.8
Katrina	3	2005	23/8/2005	30/8/2005	25/8/2005	1500	120.6
Rita	3	2005	18/9/2005	26/9/2005	24/9/2005	7	13.4
Wilma	3	2005	15/10/2005	25/10/2005	24/10/2005	5	23.5
Ike	2	2008	1/9/2008	14/9/2008	13/9/2008	20	29.9
Irma	5	2017	31/8/2017	14/9/2017	10/9/2017	134	77.16

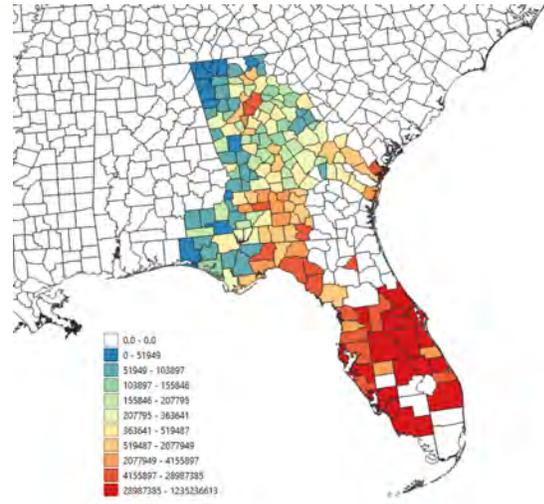
proximity, outside the disaster area. These are categorized as adjacent counties. Proximity is determined using the average latitude and longitude of urban centers within each county. Counties that do not fall into these groups are considered non-impacted. For example, Figure 5.3(a) illustrates Hurricane Irma’s trajectory and the three categories of areas, while 5.3(b) shows the SHELDDUS recorded economic losses by each county.

Consequently, firms are identified as impacted, adjacent, or non-impacted based on the classification of the county where their headquarters are located. The geographic distribution of firm headquarters at the firm-year level within our sample is illustrated in the Figure 5.4.

Table 5.3 presents the summary statistics for the variables utilized in the regression analyses of this chapter. The sample is divided into three categories based on their proximity to the event: impacted firms, adjacent firms, and non-impacted firms. Details of the definitions and calculations of the variables can be found in Table 5A.1. Several features are worth noting. Based on the statistics, firms located in the impacted regions have significantly lower ESG scores and market-to-book ratio compared to those in adjacent and non-impacted re-

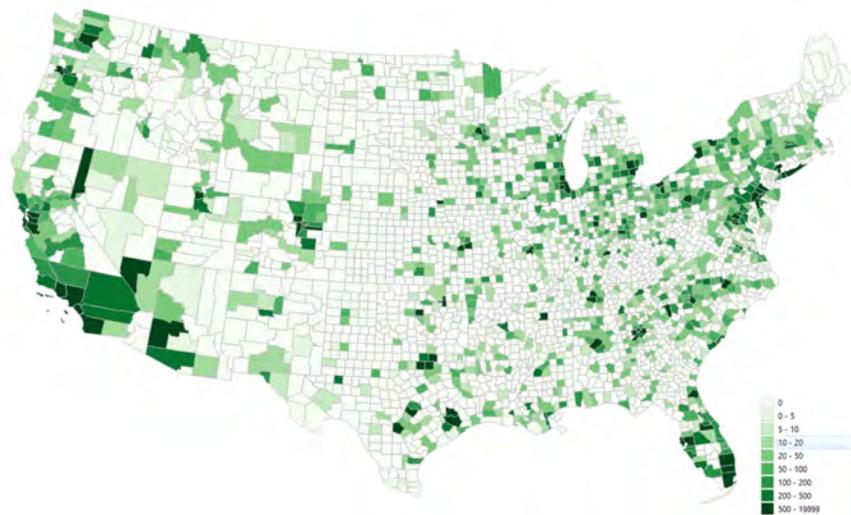


(a) Affected and Nearby Region



(b) Property loss distribution

**Figure 5.3.** Impact regions of hurricane irma.



**Figure 5.4.** Geographical distribution of Compustat firms' firm-year headquarters at the county level from 2003 to 2020.

gions. This suggests that the performance of firms disrupted by the hurricane event may lead to a reduced focus on their ESG commitments (Jiao, 2010).

### 5.3.3 Empirical Methodology

To empirically examine how the stock market reacts to the hurricane events for the green and brown firms, we report the association on CAR in response to the hurricane events for the impacted, adjacent, and non-impacted firms. Specifically, we estimate the following regression:

$$CAR_{ijst} = \beta_1 D_{is}^I \times ESG_{it} + \beta_2 D_{is}^A \times ESG_{it} + \beta_3 D_{is}^N \times ESG_{it} \\ + \beta_4 D_{is}^I + \beta_5 D_{is}^A + \beta_6 D_{is}^N + \gamma' X_{ijt} + \alpha_i + \alpha_s + \epsilon_{ijst},$$

where  $i$  indexes firms,  $j$  indexes industries,  $s$  indexes events, and  $t$  indexes years.  $ESG_{it}$  represents the one-year lagged ESG score or its composites for firm  $i$  of year  $t$ .  $D_{is}^I$ ,  $D_{is}^A$ , and  $D_{is}^N$  represent the dummy indicator for impacted, adjacent, and non-impacted firm  $i$  during the event  $s$ , respectively. The firm and event fixed effects are captured in  $\alpha_i$  and  $\alpha_s$ .  $CAR_{ijst}$  is the firm-level CAR during each event.  $X_{ijt}$  is a vector of control variables.  $\epsilon_{ijst}$  is the error term. We cluster the standard errors at the county level to account for cross-sectional correlation. Given that hurricane events represent shocks applied over time, we choose not to cluster standard errors on the time dimension.

The control variables include firm size, market-to-book ratio, leverage, gross profit, asset growth, and sales growth for each company. To mitigate the concern of forward-looking bias, these variables are lagged by one year. The computations for these variables are performed using data from the Compustat Annual and Quarterly database. Detailed methodologies for calculating each control variable are provided in Appendix 5A.1. Table 5.3 provides the summary statistics of the main and control variables utilized in our baseline results.

To reinforce the causal link suggested by our baseline findings, we employ a quasi-natural experiment approach following Cuñat et al. (2012), focusing on shareholder proposals related to ESG topics that are on the margin of approval or rejection. This analysis is carried

**TABLE 5.3**  
**Summary Statistics**

The sample encompasses 26,959 firm-event level observations spanning from 1995 to 2017. Panels A, B, and C provide summary statistics for firms impacted by hurricanes, those adjacent to the impacted areas, and firms not impacted. We sourced our ESG data from the MSCI ESG KLD database, adjusting the aggregated ESG scores to industry averages. The accounting variables for the firms were retrieved from the Compustat annual and quarterly databases. The calculations of firms fundamentals are defined in Table 5A.1.

<i>Panel A: Hurricane Impacted Firms</i>							
Variable	Mean	STD	Skewness	Kurtosis	1st Quartile	Median	3rd Quartile
Log Asset	7.264	1.585	0.333	-0.342	6.130	7.068	8.261
Market-to-Book	1.905	1.086	2.444	7.415	1.201	1.585	2.214
Leverage	0.248	0.188	0.652	0.063	0.098	0.226	0.372
Gross Profit	0.457	0.409	1.790	3.805	0.144	0.339	0.627
Asset Growth	0.212	0.329	2.113	4.979	0.023	0.108	0.272
Sale Growth	0.238	0.415	3.017	13.448	0.025	0.136	0.364
ESG Score	-1.133	3.315	-0.898	1.984	-3.074	-0.847	0.926
E Score	-0.284	1.417	-2.142	5.836	-0.335	0.169	0.220
S Score	-0.825	2.427	-0.221	0.374	-1.527	-1.110	0.491
G Score	-0.035	1.180	-0.386	0.967	0.004	0.232	0.433
<i>Panel B: Hurricane Adjacent Firms</i>							
Variable	Mean	STD	Skewness	Kurtosis	1st Quartile	Median	3rd Quartile
Log Asset	7.105	1.710	0.412	-0.108	5.963	6.831	8.284
Market-to-Book	2.011	1.267	2.772	9.331	1.241	1.639	2.289
Leverage	0.241	0.210	0.723	-0.102	0.047	0.229	0.371
Gross Profit	0.485	0.402	1.831	4.407	0.198	0.392	0.633
Asset Growth	0.165	0.364	3.370	13.913	-0.001	0.083	0.197
Sale Growth	0.180	0.307	2.763	11.023	0.025	0.108	0.252
ESG Score	-0.706	4.021	-0.268	1.401	-3.074	-0.847	0.926
E Score	-0.133	1.276	-2.122	6.482	0.119	0.169	0.220
S Score	-0.531	3.340	0.363	1.116	-2.087	-0.492	0.601
G Score	-0.025	1.321	-0.104	-0.207	-1.428	0.232	0.433
<i>Panel C: Non-impacted Firms</i>							
Variable	Mean	STD	Skewness	Kurtosis	1st Quartile	Median	3rd Quartile
Log Asset	7.155	1.736	0.556	-0.014	5.892	6.944	8.214
Market-to-Book	2.139	1.359	2.275	6.119	1.264	1.667	2.496
Leverage	0.217	0.202	1.051	1.044	0.027	0.189	0.331
Gross Profit	0.419	0.337	1.771	3.890	0.181	0.332	0.551
Asset Growth	0.159	0.320	2.892	10.948	0.007	0.082	0.201
Sale Growth	0.174	0.347	2.542	10.515	0.021	0.107	0.240
ESG Score	-0.201	3.726	0.093	1.182	-2.847	-0.806	1.153
E Score	-0.038	1.205	-2.329	8.608	0.146	0.169	0.220
S Score	-0.049	3.247	0.737	1.347	-1.527	0.204	0.890
G Score	-0.130	1.313	-0.241	0.234	-1.529	0.232	0.262

out using a Regression Discontinuity Design (RDD). Specifically, we estimate the following regression model:

$$CAR_{ijst} = \theta D_{is}^T + P_r(X_{ijt}, \gamma_r) + P_l(X_{ijt}, \gamma_l) + \alpha_i + \alpha_s + \epsilon_{ijst}, \quad (5.1)$$

where the indices of the variables and the fixed effects are the same as the baseline model.  $D_{is}^T$  is the dummy representing the treatment effect, which in the context of this chapter indicates the ESG proposal passed on shareholder meetings at the margin. The model allows different polynomials  $P_r(X_{ijt}, \gamma_r)$  and  $P_l(X_{ijt}, \gamma_l)$  on both sides of the treatment.

## 5.4 Results

In this section, we provide the results for cumulative abnormal return during hurricane events, the improvements ESG brings to climate resilience, the channel testing, and long-term firms fundamentals.

### 5.4.1 Impact of Hurricane Events on Stock Returns

In Table 5.4, we report the CARs for the full sample of firms as well as the subsamples of impacted, adjacent, and other (non-impacted) firms. Panel A shows the CARs using the window of hurricane events. For the full sample, the mean  $CAR(\text{Start}, \text{End})$ ,  $CAR(\text{Start}, \text{Landfall})$ , and  $CAR(\text{Landfall}, \text{End})$  are all positively significant at the 5% level. This positive significance is likely driven by the demand for humanitarian resources and infrastructure rebuilding created by the hurricane shocks. The subsample results show that these positive returns are mostly driven by the non-impacted firms: the mean  $CAR(\text{Start}, \text{End})$ ,  $CAR(\text{Start}, \text{Landfall})$ , and  $CAR(\text{Landfall}, \text{End})$  for these firms are all positively significant at the 5% level, with larger magnitudes and higher significance. In contrast, the mean  $CAR(\text{Start}, \text{End})$  and  $CAR(\text{Start}, \text{Landfall})$  for impacted firms are negative and significant. The mean  $CAR(\text{Start}, \text{End})$  for the adjacent firms is negative and significant at a lower magnitude than the impacted firms. The median CARs show a similar pattern. Tests for equality in

CAR(Start, End), CAR(Start, Landfall), and CAR(Landfall, End) between impacted and non-impacted subsamples are all rejected at the 5% significance level.

Following the convention in the literature, we also calculate the CARs using the arbitrary days window before and after the hurricane landfall dates. Panel B shows the CAR(-1, 1), CAR(-2, 2), CAR(-5, 5), CAR(-10, 10), and CAR(-20, 20) for the full sample and the subsamples. The mean and median CAR(-2, 2), and (-5, 5) are significantly positive for the full sample as well as for the non-impacted subsample. However, only the tests for equality in CAR(-10, 10) and CAR(-20, 20) between impacted and non-impacted subsamples are rejected at the 5% significance level.

Overall, the results presented in Table 5.4 show that impacted firms experience lower returns, while non-impacted firms see higher returns during hurricane events. Significantly, these differences are more accurately captured when calculating CARs over the hurricane event window. This approach naturally accounts for the dual uncertainties associated with hurricanes: (1) the ambiguity surrounding whether a forming hurricane will make landfall, and (2) the extent of damage and loss post-landfall. Considering these benefits, we henceforth utilize CARs calculated over the hurricane event window to assess the financial impact of hurricane shocks.

TABLE 5.4

**Univariate Analysis of Cumulative Abnormal Returns: Assessing Hurricane Impact on Impacted and Non-Impacted Firms**

This table provides a detailed univariate analysis of CAR to evaluate the impact of hurricanes on firms, segmented by their proximity to the event. Panel A utilizes event windows to analyze CAR across three periods: from the start date to the landfall date, from the landfall date to the end date, and from the start date to the end date, offering insights into market behavior at different stages of the hurricane's impact. Panel B employs fixed windows for CAR analysis, examining five intervals: (-1, 1), (-2, 2), (-5, 5), (-10, 10), and (-20, 20) days around the event, allowing for a consistent comparison across various durations. The analysis spans four categories of firms: the Full Sample, Impact Firms, Adjacent Firms, and Other Firms, with the last column highlighting the differences in CAR between Impacted and Other Firms. Each category is further divided into two sub-columns indicating the mean and median CAR, with standard errors provided in parentheses below each value. “\*\*\*” stands for  $p$ -values  $< 0.01$ , “\*\*” stands for  $p$ -values  $< 0.05$ , and “\*” stands for  $p$ -values  $< 0.1$ .

<i>Panel A: Cumulative Abnormal Return with Event Window, in %</i>											
	Full Sample		Impacted Firms (I)		Adjacent Firms (A)		Other Firms (O)		Difference: I - O		
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	
CAR(Start,End)	0.276*** (0.000)	0.276*** (0.000)	-1.741*** (0.000)	-1.741*** (0.000)	-0.319** (0.044)	-0.319*** (0.003)	0.388*** (0.000)	0.388*** (0.000)	-2.129*** (0.000)	-2.129*** (0.000)	
CAR(Start, Landfall)	0.211*** (0.000)	0.211*** (0.000)	-1.221*** (0.000)	-1.221** (0.010)	-0.336 (0.277)	-0.336 (0.224)	0.295*** (0.000)	0.295*** (0.000)	-1.517*** (0.000)	-1.517*** (0.000)	
CAR(Landfall, End)	0.069** (0.033)	0.069 (0.142)	-0.348 (0.110)	-0.348 (0.182)	-0.223 (0.142)	-0.223* (0.081)	0.098*** (0.003)	0.098** (0.032)	-0.445* (0.043)	-0.445* (0.071)	

<i>Panel B: Cumulative Abnormal Return with Fixed Window, in %</i>											
	Full Sample		Impacted Firms (I)		Adjacent Firms (A)		Other Firms (O)		Difference: I - O		
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	
CAR(-1,1)	-0.040 (0.316)	-0.080** (0.013)	-0.153 (0.496)	0.103 (0.614)	-0.179 (0.444)	-0.103 (0.335)	-0.033 (0.429)	-0.085** (0.022)	-0.120 (0.599)	0.188 (0.888)	
CAR(-2,2)	0.151*** (0.004)	0.012* (0.060)	-0.189 (0.538)	-0.173 (0.553)	-0.350 (0.348)	0.112 (0.737)	0.176*** (0.001)	0.012** (0.037)	-0.365 (0.242)	-0.186 (0.352)	
CAR(-5,5)	0.423*** (0.000)	0.242*** (0.000)	-0.258 (0.606)	-0.257 (0.661)	-0.394 (0.438)	-0.177 (0.414)	0.468*** (0.000)	0.277*** (0.000)	-0.726 (0.152)	-0.534 (0.202)	
CAR(-10,10)	-0.001 (0.991)	-0.054 (0.989)	-1.304* (0.055)	-0.839* (0.062)	-1.375* (0.056)	-1.071** (0.040)	0.079 (0.499)	-0.005 (0.485)	-1.383** (0.045)	-0.834** (0.049)	
CAR(-20,20)	0.184 (0.310)	0.226 (0.110)	-3.336*** (0.003)	-1.912*** (0.009)	-3.384*** (0.006)	-2.176*** (0.003)	0.395** (0.033)	0.316*** (0.008)	-3.732*** (0.001)	-2.228*** (0.002)	

## 5.4.2 ESG and Climate Resilience

### Univariate Test

In Table 5.5, we report the CARs for the three subsamples of impacted, adjacent, and other (non-impacted) firms in Panels A to C. Within each subsample, we further divided the firms into high and low ESG groups according to the sample median of adjusted ESG scores. Panel A shows that the differences in mean  $CAR(\text{Start}, \text{End})$  and  $CAR(\text{Start}, \text{Landfall})$  between the high and low ESG groups among the impacted firms are positive and significant at the 5% level. Panel B reveals that the difference in mean  $CAR(\text{Start}, \text{End})$  between the high and low ESG groups among adjacent firms is positive and significant at the 5% level, albeit at a lower magnitude than in the impacted firms. Panel C shows that the differences in mean  $CAR(\text{Start}, \text{End})$  and  $CAR(\text{Landfall}, \text{End})$  between the high and low ESG groups are negative and significant at the 5% level.

The negative significance among the non-impacted firms is probably driven by the substitution effect: the low (high) ESG groups in the impacted region are more (less) affected by the hurricane shock, thus their peers in the same industry but located at the non-impacted counties would benefit more (less) by providing substitution products or services to occupy the supply gap left by the business disruption of the impacted firms.

Results in Table 5.5 provide evidence that a good ESG status improves resilience during hurricane events for impacted firms. However, for non-impacted firms, a good ESG status may result in them benefiting less through the substitution effect during hurricane events.

### Cross-sectional Regression Analysis

To better estimate the cross-sectional variation in firms' climate resilience, we present estimates from the multivariate regression using  $CAR(\text{Start}, \text{End})$  as the dependent variable and the interaction between firms' hurricane exposure dummy and adjusted ESG scores as the main variables. The OLS model utilized are as described in Section 5.3.3.

The results are presented in Table 5.6.  $T$ -statistics in this table are calculated using

**TABLE 5.5**

**Univariate Analysis of Cumulative Abnormal Returns: Assessing Hurricane Impact on High and Low ESG Score Firms**

This table provides a detailed univariate analysis of CAR to evaluate the impact of hurricanes on firms, segmented by their proximity to the event. Panel A, B and C cover the hurricane-impacted firms, the hurricane-adjacent firms, and the non-impacted firms. Each panel contains three rows, presenting CAR across three periods: from the start date to the landfall date, from the landfall date to the end date, and from the start date to the end date, offering insights into market behavior at different stages of the hurricane’s impact. Each panel also contains three columns, showing the CAR on high ESG firms (H), low ESG firms (L) and the difference between the two. Each column is further divided into two sub-columns indicating the mean and median CAR, with standard errors provided in parentheses below each value to indicate variability. “\*\*\*” stands for  $p$ -values  $< 0.01$ , “\*\*” stands for  $p$ -values  $< 0.05$ , and “\*” stands for  $p$ -values  $< 0.1$ .

<i>Panel A: Hurricane-Impacted Firms</i>						
	High ESG Group (H)		Low ESG Group (L)		Difference: H - L	
	Mean	Median	Mean	Median	Mean	Median
CAR(Start,End)	-1.341 *** (0.000)	-1.341 *** (0.000)	-2.979 *** (0.000)	-2.979 *** (0.000)	1.638 *** (0.000)	1.638 * (0.056)
CAR(Start,Landfall)	-0.623 (0.259)	-0.623 (0.383)	-2.320 *** (0.000)	-2.320 *** (0.001)	1.697 ** (0.023)	1.697 (0.126)
CAR(Landfall,End)	-0.390 (0.222)	-0.390 (0.307)	-0.889 *** (0.005)	-0.889 * (0.075)	0.500 (0.265)	0.500 (0.641)
<i>Panel B: Hurricane-Adjacent Firms</i>						
	High ESG Group (H)		Low ESG Group (L)		Difference: H - L	
	Mean	Median	Mean	Median	Mean	Median
CAR(Start,End)	-0.298 (0.185)	-0.298 (0.331)	-1.166 *** (0.000)	-1.166 *** (0.000)	0.869 ** (0.014)	0.869 *** (0.002)
CAR(Start,Landfall)	-0.479 (0.289)	-0.479 (0.478)	-0.818 (0.118)	-0.818 * (0.089)	0.339 (0.623)	0.339 (0.391)
CAR(Landfall,End)	-0.068 (0.790)	-0.068 (0.528)	-0.571 ** (0.021)	-0.571 ** (0.017)	0.503 (0.154)	0.503 (0.202)
<i>Panel C: Non-Impacted Firms</i>						
	High ESG Group (H)		Low ESG Group (L)		Difference: H - L	
	Mean	Median	Mean	Median	Mean	Median
CAR(Start,End)	0.193 *** (0.000)	0.193 *** (0.000)	0.380 *** (0.000)	0.380 *** (0.000)	-0.187 *** (0.002)	-0.187 ** (0.027)
CAR(Start,Landfall)	0.094 (0.225)	0.094 (0.117)	0.229 ** (0.010)	0.229 ** (0.018)	-0.136 (0.250)	-0.136 (0.463)
CAR(Landfall,End)	0.014 (0.792)	0.014 * (0.065)	0.182 *** (0.000)	0.182 *** (0.008)	-0.168 ** (0.022)	-0.168 *** (0.001)

standard errors adjusted for heteroskedasticity and county clustering. In Column 1, we report the DID results where the firm’s ESG status is measured by the complete ESG scores. We find that the coefficient estimated on the interaction between the impacted dummy and ESG scores is positive and significant at the 1% level, while the coefficient estimated on the impacted dummy is negative and significant at the 1% level. This result suggests that the green firms are more resilient towards the disruptive hurricane events than the brown firms.

Next, in Columns 2 through 4, we reestimate the DID analysis from Column 1 but substitute the ESG rating into E, S, and G ratings (as defined in Section 5.3.2). Column 3 shows that the interaction between the impacted dummy and S scores is positive and significant at the 1% level, and the magnitude of the coefficient is twice as large as the coefficient estimated in Column 1. Conversely, the coefficient estimated on the interaction between the impacted dummy and E and G scores is not statistically significant, as reported in columns 2 and 4. The coefficient estimated on the impacted dummy is negative and significant at the 1% level throughout the three columns. These results suggest that the climate resilience improvement of the green assets is mainly sourced from its social components, rather than its ‘green’ (environmental) components. This is unfavourable to the climate risk hedging view of the improvement, as the firms’ climate risk hedging status is directly captured in E scores.

Overall, the regression results reported in Table 5.6 confirm the univariate results reported in Table 5.5 and support the market sentiment and financial stability proxy views.

### 5.4.3 Natural and Unnatural ESG Firms

In this section, we further support the market sentiment view by repeating the baseline results, substituting the individual ESG scores with the natural ESG indicator. These indicators represent an alternative approach to assessing firms’ ESG standing, not through their individual scores but via the average ESG score of their respective industries. Specifically, we calculate the average excess ESG status and the distinct ESG components for firms across the Fama-French 48 industries. We then categorize our sample into natural and unnatural

**TABLE 5.6**

**Baseline Analysis: The Influence of ESG Status and Hurricane Impact on Firms**

This table presents the results of regression analyses aimed at exploring the relationship between firms' Environmental, Social, and Governance (ESG) statuses and their varying degrees of impact from hurricanes. Specifically, columns (1) through (4) detail the regression outcomes associated with each component of ESG (namely E, S, G, and the composite ESG score) and their interactions with firms' hurricane impact statuses, categorized as Impacted, Adjacent, and Others. The coefficients are meticulously quantified, with corresponding t-statistics provided in parentheses directly below each coefficient. “\*\*\*” stands for  $p$ -values  $< 0.01$ , “\*\*” stands for  $p$ -values  $< 0.05$ , and “\*” stands for  $p$ -values  $< 0.1$ .

	(ESG):=(ESG) (1)	(ESG):=(E) (2)	(ESG):=(S) (3)	(ESG):=(G) (4)
<i>Main Variables</i>				
Impacted $\times$ (ESG)	0.00213*** (2.73058)	-0.00055 (-0.29806)	0.00491*** (2.82163)	-0.00487 (-1.15023)
Adjacent $\times$ (ESG)	0.00027 (0.32764)	0.00121 (0.42302)	0.00079 (0.82689)	-0.00335 (-1.12528)
Others $\times$ (ESG)	-0.00022 (-0.80891)	-0.00017 (-0.27631)	0.00001 (0.03363)	-0.00129 (-1.41294)
Impacted Region	-0.01496*** (-3.54831)	-0.01797*** (-3.42427)	-0.01388*** (-4.38039)	-0.01848*** (-3.42251)
Adjacent Region	0.00071 (0.22214)	0.00012 (0.03460)	0.00069 (0.21744)	-0.00050 (-0.14920)
<i>Control Variables</i>				
Log(Lagged Asset)	-0.00343 (-1.31933)	-0.00356 (-1.32184)	-0.00354 (-1.32847)	-0.00393 (-1.41019)
Lagged Market-to-Book	-0.00011 (-0.11763)	-0.00010 (-0.10639)	-0.00012 (-0.12801)	-0.00011 (-0.11810)
Lagged Leverage	-0.01695 (-1.49373)	-0.01679 (-1.51159)	-0.01675 (-1.48902)	-0.01667 (-1.48334)
Lagged Gross Profit	-0.01613* (-1.70273)	-0.01749* (-1.72982)	-0.01560* (-1.75088)	-0.01786* (-1.75716)
Lagged Asset Growth	0.00045 (0.54729)	0.00045 (0.55171)	0.00045 (0.54269)	0.00048 (0.59769)
Lagged Sale Growth	0.00001 (1.28866)	0.00001 (1.26666)	0.00001 (1.36418)	0.00001 (1.28979)
Constant	0.03472 (1.55744)	0.03624 (1.55351)	0.03526 (1.57090)	0.03883 (1.61450)
Observations	19,813	19,813	19,813	19,813
R-squared	0.223	0.223	0.224	0.223
Firm FE	Yes	Yes	Yes	Yes
Year-Quarter FE	Yes	Yes	Yes	Yes
County SE	Yes	Yes	Yes	Yes

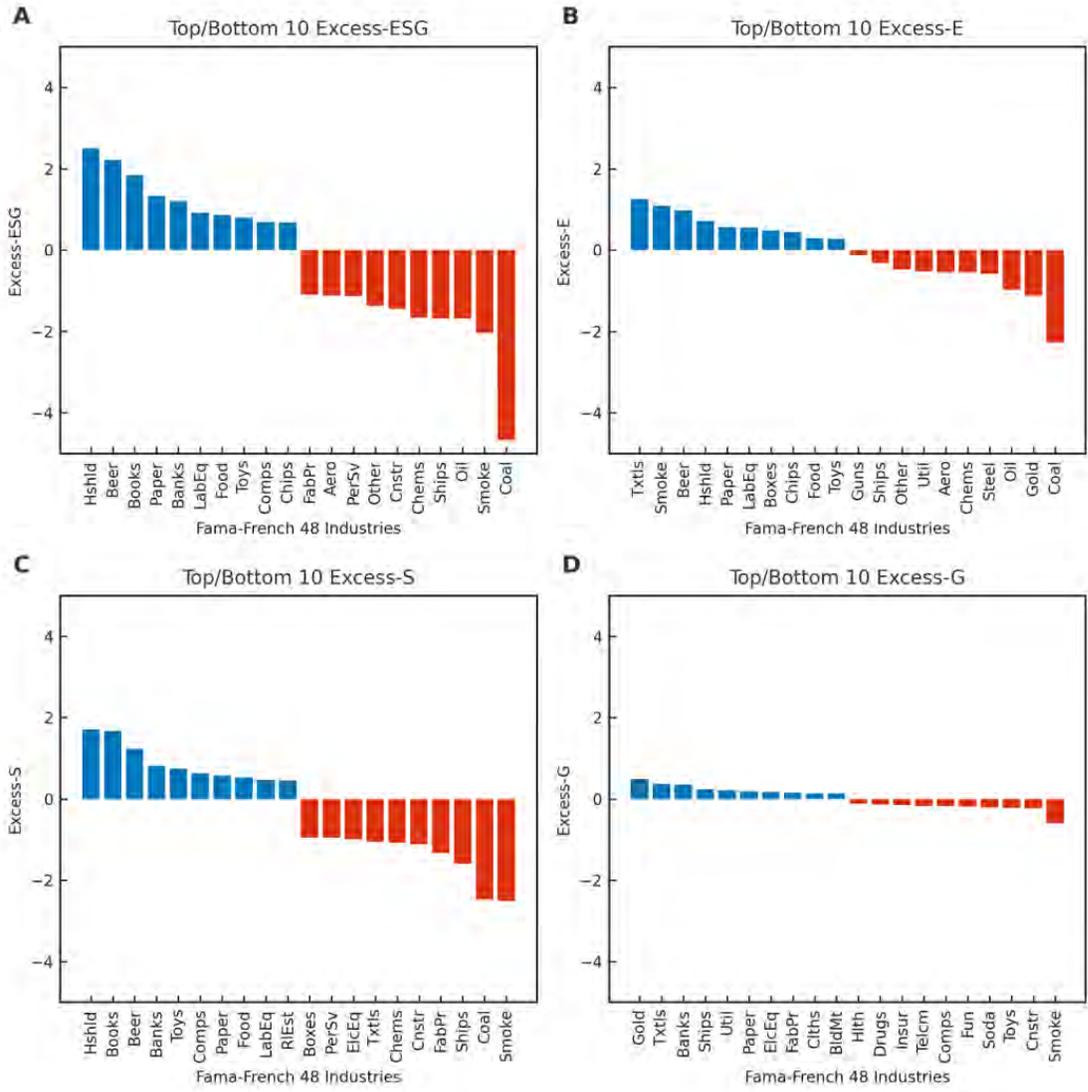
ESG firms based on the average excess ESG of the firms' industries.

The top and bottom 10 industries by ESG and each component are illustrated in Figure 5.5. Some industry rankings are generally consistent across four classifications, such as consumer goods (Hshld), beer & liquor (Beer), and measuring and control equipment (LabEq), which are generally considered strong in natural ESG, E, and S sectors. Additionally, the coal industry (Coal) is generally regarded as one of the most unnatural industry by ESG, E, and S standards. However, the decomposition of ESG scores is not merely a proxy for the composite ESG scores. The designation of an industry's ESG status as "natural" or "unnatural" changes depending on which component of ESG is used for measurement. For instance, tobacco products (Smoke) are considered the second most natural E industry due to the limited pollution generated during production, but they are viewed as one of the most unnatural S industries due to their adverse societal impacts. The logic for G industries is distinct from the rest of the measures. Top G industries such as precious metals (Gold), textiles (Txtls), banking (Banks), shipbuilding & railroad equipment (Ships), and utilities (Util) are all considered naturally heavily regulated or competitive.

Should firms within the "natural" ESG category exhibit heightened resilience to hurricane impacts, it would suggest a preference by investors for the inherent characteristics of a firm's industry over its specific ESG performance. This result would support the behavior and market sentiment channel. Nonetheless, while this observation offers insights, it does not entirely negate the influence of other channels. A more thorough examination of this channel would require an analysis of trading activities, as elaborated in Section 5.2.1.

Table 5.7 reports on the climate resilience of natural and unnatural ESG firms during hurricane events. Column 1 shows that the coefficient associated with the interaction between the impacted indicator and the natural ESG indicator is positive and significant at the 5% level. This result suggests that an industry with an average outstanding ESG status is more climate resilient towards hurricane events, which, as established, favors the market sentiment view.

In Columns 2 to 4, we replace the aggregate ESG indicator with separate natural indicators for Environmental (E), Social (S), and Governance (G) and then re-estimate the



**Figure 5.5.** Natural and unnatural ESG industries, classified by their excess ESG scores. The industries are sorted into top and bottom 10 categories based on their Excess ESG Scores. (A-D) illustrate the distribution of Excess Scores, distinguishing between Natural and Unnatural ESG industries. These classifications are made according to their respective excess ESG, E, S, and G Scores.

**TABLE 5.7**

**Natural and Unnatural ESG Firms' Resilience to Hurricane Impact**

This table presents the results of regression analyses aimed at exploring the relationship between the natural and unnatural ESG firms' and their resilience towards hurricanes impact. Specifically, columns (1) through (4) detail the regression outcomes associated with each component of ESG (namely E, S, G, and the composite ESG score) and their interactions with firms' hurricane impact statuses, categorized as Impacted, Adjacent, and Others. The coefficients are meticulously quantified, with corresponding t-statistics provided in parentheses directly below each coefficient. “\*\*\*” stands for  $p$ -values  $< 0.01$ , “\*\*” stands for  $p$ -values  $< 0.05$ , and “\*” stands for  $p$ -values  $< 0.1$ .

	(ESG):=(ESG) (1)	(ESG):=(E) (2)	(ESG):=(S) (3)	(ESG):=(G) (4)
<i>Main Variables</i>				
Impacted $\times$ (Natural ESG)	0.01694** (2.23631)	0.01694** (2.23631)	0.01820* (1.95534)	-0.00934 (-0.83835)
Adjacent $\times$ (Natural ESG)	0.00939 (1.17106)	0.00939 (1.17106)	0.01775* (1.67462)	-0.00295 (-0.40344)
Impacted Region	-0.02448*** (-5.77006)	-0.02448*** (-5.77006)	-0.02666*** (-4.49586)	-0.01350*** (-2.61018)
Adjacent Region	-0.00466 (-0.84849)	-0.00466 (-0.84849)	-0.01069 (-1.35857)	0.00122 (0.28727)
Observations	19,813	19,813	19,813	19,813
R-squared	0.223	0.223	0.223	0.223
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Year-Quarter FE	Yes	Yes	Yes	Yes
County SE	Yes	Yes	Yes	Yes

coefficients. Column 3 reveals that the interaction between the impacted indicator and the natural S indicator is positively correlated and statistically significant at 10%. This finding aligns with the results presented in Table 5.6, albeit with a lower magnitude and statistical significance. Moreover, the interaction between the adjacent indicator and the natural S indicator is also positive and significant at 10%. This observation is consistent with Dessaint and Matray (2017) such that investors overreact to salient risk. Column 2 demonstrates that the interaction between the impacted indicator and the natural E indicator is positive and significant at 5%. This outcome suggests that industries with superior environmental practices exhibit greater resilience to hurricane impacts, a relationship that does not extend to individual firms.

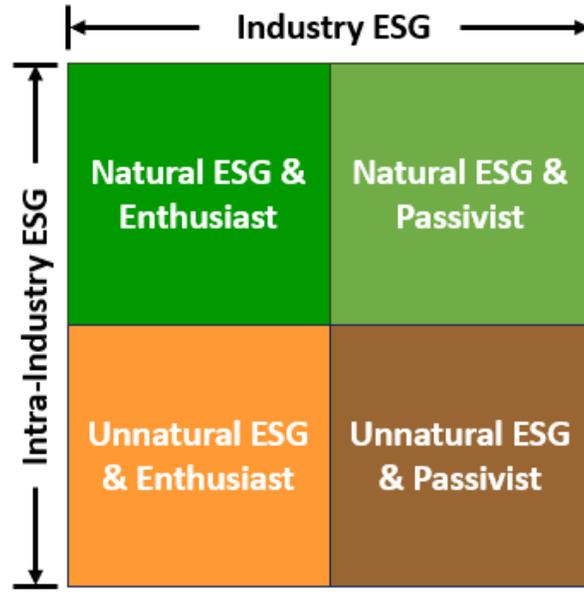
In summary, Table 5.7 demonstrates that the baseline results are preserved when replacing the firm-specific ESG measure with natural and unnatural ESG indicators defined at the industry level. These results support the market sentiment view, suggesting that investors favor firms within industries recognized for their ESG, despite some ambiguity regarding their specific ESG statuses. We will further elaborate on this mechanism in the subsequent section.

#### 5.4.4 The Impact of Being ESG Enthusiasts

As outlined in Section 5.4.3, the market sentiment view suggests that investors may not clearly distinguish among firms during their selection process, often favoring industries recognized for natural ESG rather than conducting thorough evaluations of each firm's ESG performance. In this section, we provide additional support for the market sentiment hypothesis by demonstrating that a high ESG rating within industries not typically recognized for ESG does not necessarily enhance a firm's resilience to climate-related challenges.

We calculate an industry-adjusted ESG score for each firm, categorizing them as high ESG performers or low ESG performers based on whether they score above or below the industry median, respectively.

If being categorized as an ESG enthusiast is associated with improved climate resilience, it supports the climate hedging and financial stability proxy measure channels. Specifically,



**Figure 5.6.** ESG enthusiasts in natural and unnatural ESG industries.

firms labeled as ESG enthusiasts are anticipated to demonstrate superior financial health and a proactive stance towards managing climate risks, thereby bolstering their resilience to climate-induced adversities. Conversely, should the status of a firm as an ESG enthusiast not be associated with enhanced climate resilience, this would underscore the market sentiment hypothesis. It suggests that investors’ selection processes might not comprehensively evaluate individual firms’ ESG performances but rather heavily rely on impressions of the corresponding industries, aligning with the view that there is a blurred distinction among firms in their investment decisions.

To conduct this analysis, we utilize a Difference-in-Differences-in-Differences (DDD) model to establish the relationship between a firm’s ESG enthusiasm and the cumulative abnormal return (CAR) it experiences during hurricane events. We interact three key variables in our model: the impact status indicator (categorized as impacted, adjacent, or other), the non-natural ESG industry indicator, and the ESG enthusiasm indicator, which is our primary variable of interest. Unlike in Table 5.7, where we used a natural ESG indicator, here we switch to a non-natural ESG indicator—labeling non-natural ESG firms as 1 and natural ESG firms as 0. We hypothesize that the effects of ESG enthusiasm will be more pronounced within non-natural ESG industries. Should the ESG efforts of these firms be

reflected in their CAR during hurricane events, it would substantiate the claim that investors genuinely value ESG enthusiasm. Conversely, a lack of correlation would suggest that investors may struggle to discern ESG commitments at the firm level, potentially indicating a lack of clarity in their ESG selection process.

Column 1 of Table 5.8 presents the effects of ESG enthusiasm on firms' climate resilience during hurricane events. The results show that firms identified as enthusiasts within non-natural ESG industries do not enjoy any benefits in terms of CAR during hurricanes, irrespective of their location in impacted, adjacent, or non-impacted (others) counties. Importantly, the coefficient for the "others" indicator combined with ESG enthusiasm is significantly negative at the 1% level. This indicates that being an ESG enthusiast, particularly within non-natural ESG industries, is associated with adverse effects on CAR for firms situated in regions unaffected by hurricanes.

Columns 2 to 4 of Table 5.8 substitute the general ESG indicator and ESG enthusiast indicator with specific E, S, and G indicators, along with their respective enthusiast indicators, to replicate the analysis conducted in Column 1. In alignment with the findings from Column 1, the data show that firms identified as enthusiasts in the non-natural E, S, or G sectors do not exhibit any statistically enhanced climate resilience.

Overall, Table 5.8 indicates that firms identified as enthusiasts within non-natural ESG industries do not experience additional climate resilience. These findings lend support to the market sentiment view, suggesting that the extra efforts these firms invest in ESG improvements, compared to their industry peers, are not recognized or rewarded during hurricane events.

#### **5.4.5 Long-term ESG Status and Fundamentals after Hurricane Shocks**

In this section, we explore the potential impact of hurricane shocks on firms' long-term ESG metrics and fundamental financial indicators, including asset growth, sales growth, gross profit, and leverage. To prevent forward-looking bias, we use ESG scores lagged by

**TABLE 5.8**  
**ESG Enthusiasts' Resilience to Hurricane Impact**

This table presents the results of regression analyses aimed at exploring the relationship between the firms' ESG enthusiasm and their resilience towards hurricanes impact. Specifically, columns (1) through (4) detail the regression outcomes associated with each component of ESG (namely E, S, G, and the composite ESG score) and their interactions with firms' hurricane impact statuses, categorized as Impacted, Adjacent, and Others. The coefficients are meticulously quantified, with corresponding t-statistics provided in parentheses directly below each coefficient. “\*\*\*” stands for  $p$ -values  $< 0.01$ , “\*\*” stands for  $p$ -values  $< 0.05$ , and “\*” stands for  $p$ -values  $< 0.1$ .

	(ESG):=(ESG) (1)	(ESG):=(E) (2)	(ESG):=(S) (3)	(ESG):=(G) (4)
<i>Main Variables</i>				
Impacted $\times$ (Non-Natural ESG) $\times$ Enthusiast	-0.01348 (-0.80382)	0.00151 (0.10831)	-0.00406 (-0.40373)	0.00028 (0.01935)
Adjacent $\times$ (Non-Natural ESG) $\times$ Enthusiast	-0.00376 (-0.19011)	0.00200 (0.10325)	-0.00742 (-0.43728)	-0.00259 (-0.14205)
Others $\times$ (Non-Natural ESG) $\times$ Enthusiast	0.00359 (0.74541)	-0.00254 (-0.59529)	-0.00208 (-0.47851)	-0.00194 (-0.43933)
Impacted $\times$ Enthusiast	0.01340 (1.44386)	-0.02783 (-1.40384)	0.00736 (0.52270)	-0.00306 (-0.19243)
Adjacent $\times$ Enthusiast	0.00020 (0.01872)	-0.03887* (-1.79923)	0.00563 (0.72890)	-0.00364 (-0.40853)
Others $\times$ Enthusiast	-0.00615*** (-2.76758)	-0.00092 (-0.30679)	-0.00113 (-0.45088)	-0.00273 (-1.16693)
Impacted $\times$ (Non-Natural ESG)	-0.02393*** (-2.58905)	-0.02616*** (-3.10508)	-0.02576*** (-3.11237)	-0.02641*** (-2.78461)
Adjacent $\times$ (Non-Natural ESG)	-0.01345* (-1.79387)	-0.01467** (-2.13301)	-0.01366* (-1.90176)	-0.01459** (-1.98567)
Impacted	-0.00338 (-0.64023)	0.00012 (0.02321)	-0.00180 (-0.35231)	-0.00089 (-0.16521)
Adjacent	0.00603 (1.44163)	0.00824** (2.17123)	0.00578 (1.44982)	0.00707 (1.62738)
Observations	25,549	25,549	25,549	25,549
R-squared	0.209	0.209	0.209	0.209
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Year-Quarter FE	Yes	Yes	Yes	Yes
County SE	Yes	Yes	Yes	Yes

one year and predict the outcomes of these variables one year post-hurricane. This approach enables us to examine whether the financial stability proxy of ESG status is a significant channel that affects firms' climate resilience.

Panel A of Table 5.9 presents the findings. Columns 1-4 detail the impacts of hurricane shocks and the interaction between the impact indicator and firms' ESG status on their long-term ESG performance, including the individual components of ESG. While the coefficients for the hurricane impact indicator alone do not reach statistical significance, the interaction between the impact indicator and the firms' ESG scores is positively significant for both overall ESG and E scores as dependent variables at the 1% level. This suggests that firms with strong ESG credentials are more aware of climate risks; hurricane shocks heighten their concern, which in turn leads to increased commitment to climate risk mitigation and adaptation strategies. Columns 5-8 detail the effects of hurricane shocks and the interaction between the impact indicator and firms' ESG status on their long-term financial fundamentals. The analysis reveals that hurricane shocks do not significantly impact the long-term financial metrics of firms, consistent with previous findings (Alok et al., 2020).

TABLE 5.9

**Impact of Firms' ESG Status and Hurricane Shocks on Long-term ESG Status and Fundamentals**

This table presents the regression results on how the interaction between firms' ESG status in the prior year (year t-1) and their exposure to hurricane impacts in the current year (year t) influences their subsequent ESG status and financial fundamentals in the following year (year t+1). The analysis is structured into two panels for a comprehensive examination of these dynamics. In Panel A, the main variable is the interaction of firms' initial ESG status and hurricane impact. In Panel B, we focus specifically on the E and S components. The table is organized into eight columns: Columns (1) to (4) detail the dependent variables related to ESG status in year t+1, specifically ESG, E, S, and G components, respectively. Columns (5) to (8) shift the focus to firms' financial fundamentals in year t+1, encompassing Asset Growth, Sales Growth, Gross Profit, and Leverage. The control variables, Firm and Year-Quarter Fixed Effects are included. Standard error are clustered at county level. Corresponding t-statistics are provided in parentheses directly below each coefficient. "\*\*\*\*" stands for p-values < 0.01, "\*\*\*" stands for p-values < 0.05, and "\*\*" stands for p-values < 0.1.

<i>Panel A: Hurricane Events and Initial ESG Status</i>								
	ESG [t+1]	E [t+1]	S [t+1]	G [t+1]	Asset Growth [t+1]	Sale Growth [t+1]	Gross Profit [t+1]	Leverage [t+1]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Main Variables</i>								
Impacted in Year t (Impacted)	0.17561 (0.58826)	0.21162 (1.59368)	-0.08958 (-0.44837)	-0.08958 (-0.44837)	-0.01268 (-0.22479)	0.21203 (1.14404)	-0.01853 (-0.26904)	0.00212 (0.20706)
(Impacted) × ESG [t-1]	0.19625*** (3.14762)	0.04119** (2.36469)	0.05930 (1.42734)	0.05930 (1.42734)	-0.00910 (-0.72547)	-0.01784 (-1.01912)	0.01265 (1.20973)	-0.00128 (-0.62669)
Observations	9,239	9,239	9,239	9,239	9,594	9,588	9,591	9,575
R-squared	0.749	0.732	0.775	0.775	0.128	0.448	0.926	0.751
<i>Panel B: Hurricane Events and Initial E and S Status</i>								
	ESG [t+1]	E [t+1]	S [t+1]	G [t+1]	Asset Growth [t+1]	Sale Growth [t+1]	Gross Profit [t+1]	Leverage [t+1]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Main Variables</i>								
Impacted in Year t (Impacted)	0.12955 (0.41054)	0.10734 (0.97121)	-0.07765 (-0.36481)	-0.07765 (-0.36481)	-0.01577 (-0.26136)	0.19182 (1.06809)	-0.01336 (-0.19026)	0.00226 (0.23099)
(Impacted) × E [t-1]	0.53889*** (3.49696)	0.47592*** (5.07920)	0.06847 (0.42189)	0.06847 (0.42189)	-0.00158 (-0.07683)	0.04836 (0.81171)	-0.00100 (-0.04512)	-0.00301 (-0.48120)
(Impacted) × S [t-1]	0.10931 (0.94688)	-0.02632 (-0.49120)	0.06650* (1.86986)	0.06650* (1.86986)	-0.00285 (-0.25385)	-0.02268* (-1.68158)	0.00481 (0.37466)	-0.00147 (-0.68277)
Observations	9,239	9,239	9,239	9,239	9,594	9,588	9,591	9,575
R-squared	0.749	0.732	0.775	0.775	0.128	0.448	0.926	0.751

Panel B of Table 5.9 repeats the analysis from Panel A, focusing on decomposing the ESG score into E and S scores, while omitting the G score due to its relative lack of importance in previous analyses. Consistent with the results from Panel A, we observed no significant long-term impact of hurricane shocks on firms' ESG status or financial fundamentals. However, the interaction between the impact indicator and the E score is positively significant at a 1% level, influencing both the long-term ESG score and the E score. Similarly, the interaction with the S score is positively significant at a 5% level for the long-term S score.

In summary, our analysis reveals no evidence that firms with relatively strong ESG standings gain long-term fundamental benefits following hurricane events. Consequently, we find no support for the hypothesis that the financial stability proxy channel explains the superior climate resilience observed in ESG-oriented firms during such events.

## 5.5 Robustness tests

To ensure the robustness of our findings, we conducted several additional tests. Section 5.5.1 confirms that our results are not influenced by ESG endogeneity. Section 5.5.2 strengthens our understanding of the causal relationship between ESG scores and firms' climate resilience through repeating the baseline analyses on the most impacted industries subsample.

### 5.5.1 ESG Endogeneity

To address potential endogeneity in firms' ESG status, such as reverse causality and the high correlation of ESG scores with omitted variables that may causally influence firms' climate resilience, we conducted a Regression Discontinuity Design (RDD) analysis based on our baseline results. The methodology and settings for this test are detailed in Section 5.3.3. Given the relatively small size of the shareholders' voting data, we expanded our sample to include a broader range of events. Specifically, we extended our analysis beyond the hurricanes causing over \$5 billion in losses, as listed in Table 5.2, to include NatCat events

with loss larger than \$500 million for a single state, recorded in the SHELDUS 20.0 database after 2006. This expansion ensures a sufficient dataset for RDD analyses.

Table 5.10 presents the results. We test the treatment effects of the passage of ESG proposals on firms' CAR when firms are affected during NatCat events. This analysis was conducted on the full sample and five subsamples, where shareholders' ESG proposals were passed by narrow margins, specifically within the vote percentage ranges of [-15%, 15%], [-10%, 10%], [-5%, 5%], [-2.5%, 2.5%], and [-1%, 1%]. Across the full sample and all subsamples, the interaction between the impacted indicator (NatCat) and ESG proposal passing indicator (ESG) shows a positive and statistically significant effect, with a  $p$ -value of less than 0.01.

**TABLE 5.10**  
**Regression Discontinuity Analysis: The Impact of ESG Policy Adoption on Firm Resilience to Natural Catastrophe Shocks**

This table presents the findings from a Regression Discontinuity Design analysis examining shareholder voting behavior on ESG related actions, particularly in the context of passing thresholds and their interaction with natural catastrophe shocks. The analysis leverages data from the historical catastrophe events with loss larger than 500 million 2019 dollar listed in the SHELDUS 20.0 database post-2006 to define natural catastrophe shocks. The main variable is the interaction between ESG action passing thresholds and these natural catastrophe events, aiming to uncover the impact of ESG action on firms' catastrophe resilience. Column (1) showcases the regression results for all voting outcomes, while columns (2) through (6) present differentiated findings for votes within varying margins of the threshold: +/-15%, +/-10%, +/-5%, +/-2.5%, and +/-1%, respectively. The corresponding t-statistics are provided in parentheses below each coefficient. “\*\*\*” stands for  $p$ -values < 0.01, “\*\*” stands for  $p$ -values < 0.05, and “\*” stands for  $p$ -values < 0.1.

	All (1)	+/-15% (2)	+/-10% (3)	+/-5% (4)	+/-2.5% (5)	+/-1% (6)
<i>Main Variables</i>						
ESG × NatCat	0.00933 *** (3.19708)	0.01180 *** (3.25732)	0.01156 *** (3.31880)	0.01825 *** (3.67624)	0.03569 *** (5.62645)	0.05863 *** (15.95542)
Observations	91,730	21,264	11,681	5,374	2,397	518
R-squared	0.069	0.0470	0.052	0.039	0.049	0.154
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Event FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
State SE	Yes	Yes	Yes	Yes	Yes	Yes

The ESG proposal passed on shareholders' meeting at small margin can be treated as a random treatment. Therefore, the quasi-experimental framework established via RDD analysis substantiates the causal relationship between firms' ESG initiatives and their climate resilience. We find no evidence that the observed improvement in climate resilience associated

with firms' ESG status can be attributed to the potential endogeneity of ESG scores.

## 5.5.2 Climate Resilience in the Most Impacted Industries

We further examine causality by testing whether the impact of hurricane shocks on firms is more pronounced within industries that suffer the greatest damage from hurricane events. Should causality be confirmed, it would imply that within the most impacted industries, green firms exhibit greater resilience to hurricane impacts compared to those in less affected industries.

For each hurricane event, we assess the total market capitalization loss for firms impacted by and adjacent to the event, categorized by the Fama-French 48 industries. The top five industries most affected by each hurricane are identified and presented in Table 5.11. These top five industries are subsequently aggregated and analyzed as the most affected industries in our subsequent analyses.

**TABLE 5.11**  
**Top Five Industries Most Impacted by Hurricane Strikes**

This table catalogs the impact of hurricanes in-use on market capitalization of different industries, ranking industries according to the extent of their financial losses during the hurricane event window. Column 1 lists the names of the hurricanes, serving as the pivotal events under study. Columns 2 to 6, detail the industries that experienced the most significant market capitalization losses, ranked from the first to the fifth most affected.

Hurricane (1)	Industry Market Cap Loss Rank				
	1st (2)	2nd (3)	3rd (4)	4th (5)	5th (6)
Alison	Whlsl	Util	Oil	PerSv	Cnstr
Isabel	LabEq	Mach	Mines	Banks	Drugs
Charley	Rtail	Trans	Whlsl	LabEq	Boxes
Frances	Rtail	Hlth	Chips	LabEq	Drugs
Ivan	Util	Paper	Food	Rtail	Banks
Jeanne	Paper	Drugs	Insur	RlEst	Fun
Katrina	Trans	Mines	Meals	Food	Rtail
Rita	Whlsl	Fin	Hlth	PerSv	Insur
Wilma	Hlth	Aero	Rtail	Chips	Telcm
Ike	Oil	Util	Mach	BusSv	BldMt
Irma	BusSv	Chips	Whlsl	Banks	Fin

Panel A of Table 5.12 replicates the baseline analysis from Table 5.6 using the subsample of the most affected industries. Columns 1 and 3 reveal that the interactions between the

impact indicator and both the ESG and S scores are significantly positive at the 1% and 5% levels, respectively, with magnitudes exceeding those of the corresponding coefficients in Table 5.6. Notably, Column 2 demonstrates that firms with high Environmental (E) scores report higher CAR, regardless of their positioning in impacted, adjacent, or non-impacted (other) areas. This evidence suggests that the climate risk hedging channel has explanatory power among the most impacted industries for each hurricane event.

We observed that the wholesale (Whlsl) and retail (Rtail) industries are prominently featured in Table 5.11, appearing five and four times, respectively, across various hurricanes. To validate the robustness of our findings in Panel A of Table 5.12, we excluded these two industries from our subsample to create a new subset and conducted the analyses again. The outcomes, presented in Panel B of Table 5.12, align closely with those in Panel A, exhibiting comparable magnitudes and levels of statistical significance for the coefficients.

These results demonstrate that the effect of ESG status on enhancing climate resilience is more pronounced within industries most affected by hurricane events. Furthermore, the perspective of climate risk hedging can explain the improvement in climate resilience attributable to ESG status in the industries that are most impacted.

## 5.6 Conclusion Remarks

In this study, we explore the impact of firms' ESG status on their resilience to climate-related disruptions, specifically hurricane events. By categorizing firms based on their exposure—impacted, adjacent, and non-impacted—we present compelling evidence that those with superior ESG credentials demonstrate greater resilience to disruptive climate events.

We examine three possible channels to elucidate the observed ESG benefits on climate resilience: the market sentiment, financial stability proxy, and climate risk hedging views. The market sentiment hypothesis predicts that investors favor ESG-themed assets amidst some ambiguity, occasionally without granular analysis at the individual firm level. The financial stability proxy perspective suggests ESG status as an indicator of financial health, attributing the enhanced resilience of ESG-favorable assets to the inherent stability of financially

**TABLE 5.12**  
**Robustness Analysis: The Impact of ESG Status on Climate Resilience in the Most Impacted Industries**

This table presents the results of regression analyses aimed at exploring the relationship between firms' Environmental, Social, and Governance (ESG) statuses and their varying degrees of impact from hurricanes in the most impacted industries. Specifically, columns (1) through (4) detail the regression outcomes associated with each component of ESG (namely E, S, G, and the composite ESG score) and their interactions with firms' hurricane impact statuses, categorized as Impacted, Adjacent, and Others. The coefficients are meticulously quantified, with corresponding t-statistics provided in parentheses directly below each coefficient. “\*\*\*” stands for  $p$ -values  $< 0.01$ , “\*\*” stands for  $p$ -values  $< 0.05$ , and “\*” stands for  $p$ -values  $< 0.1$ .

<i>Panel A: Subsample Analysis on Most Affected Industry</i>				
	(ESG):=(ESG) (1)	(ESG):=(E) (2)	(ESG):=(S) (3)	(ESG):=(G) (4)
<i>Main Variables</i>				
Impacted x (ESG)	0.00323*** (2.67814)	0.02284*** (4.30677)	0.00363** (2.08880)	-0.00480 (-1.52552)
Adjacent x (ESG)	0.00077 (0.38053)	0.01019** (2.23181)	0.00112 (0.33726)	-0.01057* (-1.71337)
Others x (ESG)	0.00071 (0.93541)	0.00483** (2.05556)	0.00073 (0.97265)	-0.00140 (-0.69713)
Impacted Region	-0.03634*** (-2.77410)	-0.03670*** (-2.94191)	-0.03667*** (-2.84969)	-0.04072*** (-3.08778)
Adjacent Region	-0.00621 (-0.52711)	-0.00582 (-0.54390)	-0.00568 (-0.50213)	-0.00934 (-0.80734)
Observations	2,769	2,769	2,769	2,769
R-squared	0.548	0.550	0.548	0.547
<i>Panel B: Subsample Analysis on Most Affected Industry Exclude Rtail &amp; Whlsl</i>				
	(ESG):=(ESG) (1)	(ESG):=(E) (2)	(ESG):=(S) (3)	(ESG):=(G) (4)
<i>Main Variables</i>				
Impacted x (ESG)	0.00318** (2.50787)	0.02089*** (3.54684)	0.00360** (2.06716)	-0.00415 (-1.02295)
Adjacent x (ESG)	0.00049 (0.23483)	0.00807* (1.68794)	0.00099 (0.29596)	-0.00987 (-1.48272)
Others x (ESG)	0.00054 (0.73758)	0.00404* (1.78986)	0.00062 (0.85676)	-0.00156 (-0.84040)
Impacted Region	-0.03693*** (-2.85390)	-0.03731*** (-3.04859)	-0.03735*** (-2.92804)	-0.04147*** (-3.18536)
Adjacent Region	-0.00830 (-0.71558)	-0.00815 (-0.76159)	-0.00770 (-0.70256)	-0.01132 (-0.97970)
Observations	3,311	3,311	3,311	3,311
R-squared	0.571	0.573	0.571	0.571

sound firms. Finally, the climate risk hedging hypothesis is predicated on the assumption that firms with high ESG scores are more proactive in mitigating climate risks.

Upon adjusting for endogeneity bias, our findings indicate that firms within inherently ESG-aligned industries display increased resilience to hurricane disruptions. However, firms in industries not typically associated with ESG values do not receive similar benefits, even when allocating additional resources compared to their lower-ESG peers. This outcome supports the market sentiment hypothesis. Additionally, our examination of the long-term impact of ESG status on firm fundamentals post-hurricane reveals no substantial evidence backing the financial stability proxy hypothesis. However, in analyzing a subset of the most affected industries, the climate risk hedging approach emerges as a plausible explanation for their observed resilience.

The results in this chapter underscore the significance of a high ESG status in enhancing firm resilience to climate disruptions, primarily through the lens of market sentiment. Notably, within the most severely impacted sectors, this resilience enhancement is more distinct and is partly attributable to effective climate risk management strategies.

## 5A Appendix

### 5A.1 Variable Definitions

The variable definitions are as described in Table 5A.1.

**TABLE 5A.1**  
**Variable Definitions**

Variable	Description (variable definitions in parentheses refer to Compustat designations where appropriate)
Asset	The value of total book assets ( <i>atq</i> ) in millions.
Market-to-Book	The market value of assets (book value of assets ( <i>atq</i> ) plus market value of equity ( <i>prcc_f*csho</i> ) minus book value of equity ( <i>ceq</i> )) divided by book value of assets ( <i>atq</i> ).
Leverage	The value of long-term debt ( <i>dltt</i> ) plus debt in current liabilities ( <i>dlc</i> ) divided by book value of assets ( <i>atq</i> ).
Gross Profit	Income before extraordinary item ( <i>ib</i> ) plus depreciation and amortization ( <i>dp</i> ) divided by book value of assets ( <i>atq</i> ).
Asset Growth	The percentage change in the book value of assets ( <i>atq</i> ) compared to the same quarter of the previous year.
Sale Growth	The percentage change in sale ( <i>sale</i> ) compared to the previous year.
Impacted	An indicator variable that is set to one if the county where a firm is headquartered has recorded a hurricane or NatCat loss for a specific event, and set to zero otherwise.
Adjacent	An indicator variable that is set to one if the county where a firm is headquartered, ranks among the ten nearest counties adjacent to a county that has recorded a hurricane or NatCat loss for a specific event; it is set to zero otherwise.
Others	An indicator variable that is set to one if the Impacted and Adjacent indicators are set to zero; it is set to zero otherwise.

### 5A.2 Calculation of Cumulative Abnormal Returns

For each firm  $i$ , we calculate the abnormal return using the market model. To estimate coefficients  $\alpha_i$  and  $\beta_i$ , we apply the ordinary least squares method, using the data from 200 trading days ending 40 days before the event day (thus, for an event on day 0, the estimation period spans from day -240 to day -41), drawing on daily data from the CRSP database. Specifically, we calculate:

$$R_{it} = \alpha_i + \beta_i R_{mt} + \epsilon_{it}, \quad (5A.1)$$

where  $R_{it}$  is the return on the stock of the firm  $i$  on day  $t$ ,  $\alpha_i$  is the intercept,  $\beta_i$  is the systematic risk associated with stock  $i$ ,  $R_{mt}$  is the daily return of the equally weighted CRSP market portfolio, and  $\epsilon_{it}$  is the daily risk adjusted residual for firm  $i$ . The corresponding estimated return on the stock of firm  $i$  on day  $t$  is given by

$$\hat{R}_{it} = \alpha_i + \beta_i \times R_{mt}. \quad (5A.2)$$

We then derive the abnormal return (AR) for the stock of firm  $i$  on day  $t$  by subtracting the estimated return from the actual return:

$$AR_{it} = R_{it} - \hat{R}_{it}. \quad (5A.3)$$

The cumulative abnormal return (CAR) is computed by summing up the abnormal returns over a specified period.

### 5A.3 Natural Catastrophes Used in RDD

Table 5A.2 summarizes the state level NatCat events incorporated into the analyses in RDD analysis in Section 5.5.1. For each event, the table lists its name, state, year, start and end dates, and damages. The damages are adjusted to reflect 2019 dollar values to ensure consistency in economic impact assessment across different years.

**TABLE 5A.2**  
Natural Catastrophes Utilized for RDD: Characteristics and Impact

Name	State	Year	Start Date	End Date	Damage, \$b
Hurricane Harvey	TX	2017	17/8/2017	3/9/2017	94.0
Hurricane Katrina	LA	2005	23/8/2005	31/8/2005	64.7
Hurricane Katrina	MS	2005	23/8/2005	31/8/2005	33.2
Hurricane Sandy	NJ	2012	22/10/2012	2/11/2012	28.5
California wildfires 2018	CA	2018	1/11/2018	30/11/2018	19.6
Hurricane Ike	TX	2008	1/9/2008	15/9/2008	18.0
Hurricane Frances	FL	2004	24/8/2004	10/9/2004	16.2
Hurricane Wilma	FL	2005	15/10/2005	27/10/2005	13.7
Hurricane Laura	LA	2020	20/8/2020	29/8/2020	10.7
Louisiana floods 2016	LA	2016	12/8/2016	31/8/2016	9.6
Tropical Storm Allison	TX	2001	4/6/2001	18/6/2001	7.6

*Continued on next page*

Table 5A.2 – *Continued from previous page*

Name	State	Year	Start Date	End Date	Damage, \$b
Hurricane Katrina	LA	2005	23/8/2005	31/8/2005	6.0
The Red River flood 1997	ND	1997	1/4/1997	30/4/1997	5.9
Hurricane Floyd	NC	1999	7/9/1999	19/9/1999	5.7
Hurricane Fran	NC	1996	23/8/1996	10/9/1996	5.4
Hurricanes Marilyn and Opal	FL	1995	12/9/1995	6/10/1995	5.4
Alabama tornadoes 2011	AL	2011	24/4/2011	26/4/2011	5.1
Nisqually earthquake 2001	WA	2001	27/2/2001	1/3/2001	4.5
Hurricane Michael	FL	2018	6/10/2018	16/10/2018	4.4
Louisiana flood 1995	LA	1995	7/5/1995	10/5/1995	4.3
Hurricane Irma	FL	2017	30/8/2017	14/9/2017	4.1
Hurricane Delta	LA	2020	4/10/2020	12/10/2020	3.7
Oregon wildfires 2020	OR	2020	1/9/2020	30/9/2020	3.6
Hurricane Ivan	AL	2004	4/9/2004	25/9/2004	3.6
Joplin tornado 2011	MO	2011	22/5/2011	31/5/2011	3.6
Arizona tornado 2010	AZ	2010	4/10/2010	7/10/2010	3.4
Hurricane Rita	TX	2005	18/9/2005	26/9/2005	3.0
Hurricane Matthew	FL	2016	28/9/2016	10/10/2016	3.0
California wildfire 2003	CA	2003	1/10/2003	31/10/2003	2.7
Tennessee floods 2010	TN	2010	1/5/2010	9/5/2010	2.7
Colorado hail storm 2017	CO	2017	6/5/2017	10/5/2017	2.4
Tennessee flood 2011	TN	2011	1/5/2011	31/5/2011	2.3
Texas storm 1995	TX	1995	4/5/1995	7/5/1995	2.3
New Mexico wildfire 2000	NM	2000	1/5/2000	31/5/2000	2.3
Oklahoma tornado 2013	OK	2013	19/5/2013	21/5/2013	2.3
Hurricane Dennis	FL	2005	4/7/2005	18/7/2005	2.2
TEXAS 2019 October tornado	TX	2019	20/10/2019	22/10/2019	2.0
Hurricane Florence	NC	2018	31/8/2018	18/9/2018	2.0
Michigan flood 2014	MI	2014	9/8/2014	11/8/2014	2.0
Hurricane Gustav	LA	2008	25/8/2008	7/9/2008	2.0
Houston flooding 2016	TX	2016	15/4/2016	20/4/2016	1.9
Oklahoma tornado outbreak 1999	OK	1999	2/5/1999	10/5/1999	1.8
Multiple Hurricnaes	TX	2020	1/8/2020	31/8/2020	1.7
California wildfires 2018	CA	2018	1/7/2018	31/7/2018	1.6
Tennessee flood 2020	TN	2020	1/3/2020	31/3/2020	1.6
Texas flood 2020	TX	2016	9/3/2016	10/3/2016	1.6
Hurricane Michael	GA	2018	6/10/2018	16/10/2018	1.6
Missouri tornado 2001	MO	2001	9/4/2001	11/4/2001	1.6
Hurricane Irene	VT	2011	21/8/2011	31/8/2011	1.5
california flood 1997	CA	1997	15/12/1996	15/1/1997	1.5
2003 Alabama earthquake	AL	2003	28/4/2003	30/4/2003	1.5
Hurricane Katrina	AL	2005	23/8/2005	31/8/2005	1.4
Iowa flood 2008	IA	2008	7/6/2008	30/6/2008	1.4
Washington flood 2020	WA	2020	1/9/2020	30/9/2020	1.3
Hurricane Dolly	TX	2008	1/7/2008	31/7/2008	1.3
Texas severe storms 2012	TX	2012	1/6/2012	30/6/2012	1.3
New York flood 2006	NY	2006	26/6/2006	28/6/2006	1.3
Red River flood 1997	MN	1997	1/4/1997	30/4/1997	1.2
Hurricane Irene	NY	2011	1/9/2011	30/9/2011	1.2
Texas–Oklahoma flood and tornado 2015	TX	2015	22/10/2015	25/10/2015	1.2
Illinois tornado outbreak 2013	IL	2013	16/11/2013	18/11/2013	1.2

*Continued on next page*

Table 5A.2 – *Continued from previous page*

Name	State	Year	Start Date	End Date	Damage, \$b
Mississippi River flood 2011	MS	2011	1/5/2011	31/5/2011	1.2
California flood 1998	CA	1998	1/4/1998	30/4/1998	1.2
California wildfires 2007	CA	2007	20/10/2007	16/11/2007	1.1
Hurricane Georges	MS	1998	15/9/1998	1/10/1998	1.1
Oklahoma winter weather event 2007	OK	2007	8/12/2007	11/12/2007	1.0
Texas–Oklahoma flood and tornado 2015	TX	2015	24/5/2015	25/5/2015	1.0
Ohio Winter Storm 2008	OH	2008	7/3/2008	8/3/2008	1.0
Hurricane Irene	NJ	2011	1/8/2011	31/8/2011	1.0
January 2009 North American ice storm.	AR	2009	25/1/2009	30/1/2009	1.0
Tornado outbreak of April 3, 2012	TX	2012	2/4/2012	4/4/2012	1.0
October 1998 Central Texas floods	TX	1998	17/10/1998	18/10/1998	1.0
California Earthquake 2000	CA	2000	2/9/2000	4/9/2000	0.9
Hurricane Irene	NC	2011	1/8/2011	31/8/2011	0.9
Carson River Basin 1997	NV	1997	1/1/1997	3/1/1997	0.9
Tropical Storm Imelda	TX	2019	10/9/2019	19/9/2019	0.9
Oklahoma severe weather 2008	OK	2008	1/6/2008	10/6/2008	0.9
Texas wildfires 2011	TX	2011	1/9/2011	30/9/2011	0.9
Hurricane Ike	OH	2008	1/9/2008	15/9/2008	0.9
Iowa Tornadoes 2003	IA	2003	1/8/2003	31/8/2003	0.9
Hurricane Isabel	VA	2003	6/9/2003	20/9/2003	0.9
Texas tornado outbreak 2013	TX	2013	15/5/2013	17/5/2013	0.9
June 2008 Midwest floods	IN	2008	7/6/2008	1/7/2008	0.9
Hurricane Matthew	NC	2016	28/9/2016	10/10/2016	0.9
Louisiana thunderstorm and tornadoes 2009	LA	2009	1/5/2009	31/5/2009	0.9
Kentucky tornadoes 1998	KY	1998	15/4/1998	17/4/1998	0.9
Hurricane Floyd	NJ	1999	7/9/1999	19/9/1999	0.9
Colorado floods 2013	CO	2013	1/9/2013	30/9/2013	0.8
Hurricane Irene	FL	1999	13/10/1999	24/10/1999	0.8
Hurricane Isabel	MD	2003	6/9/2003	20/9/2003	0.8
California Earthquake 2014	CA	2014	23/8/2014	25/8/2014	0.8
Colorado wildfires 2012	CO	2012	1/6/2012	30/6/2012	0.8
Arkansas tornados 2000	AR	2000	1/12/2000	4/12/2000	0.8
Minnesota storms 1998	MN	1998	14/5/1998	16/5/1998	0.8
Hurricane Lili	LA	2002	21/9/2002	4/10/2002	0.8
Nebraska flooding and tornadoes 2001	NE	2001	10/4/2001	23/4/2001	0.8
Hurricane Georges	FL	1998	15/9/1998	10/1/1998	0.8
Kentucky flood 1997	KY	1997	1/3/1997	31/3/1997	0.7
Hurricane Isaac	LA	2012	21/8/2012	3/9/2012	0.7
2006 Mid-Atlantic United States flood	PA	2006	25/6/2006	5/7/2006	0.7
Tropical Storm Leslie 2000	FL	2000	4/10/2000	12/10/2000	0.7
1998 Kissimmee tornado outbreak	FL	1998	21/2/1998	1/3/1998	0.7
2017 California floods	CA	2017	17/2/2017	18/2/2017	0.7
2011 Mississippi River Floods	MS	2011	1/4/2011	20/6/2011	0.7
Willamette Valley flood 1996	OR	1996	20/1/1996	15/2/1996	0.7
Hurricane Isabel	NC	2003	6/9/2003	20/9/2003	0.7
Southern Wisconsin flood 2008	WI	2006	1/6/2006	30/6/2006	0.7
Southern California mudflows 2018	CA	2018	1/1/2018	31/1/2018	0.6
Hurricane Erin	FL	1995	31/7/1995	6/8/1995	0.6
California wildfire 2007	CA	2007	1/6/2007	30/6/2007	0.6
Mississippi flood 2013	MS	2013	1/3/2013	31/3/2013	0.6

*Continued on next page*

Table 5A.2 – *Continued from previous page*

Name	State	Year	Start Date	End Date	Damage, \$b
Texas tornado outbreak 2014	TX	2014	27/4/2014	30/4/2014	0.6
Florida wildfires 1998	FL	1998	1/7/1998	31/7/1998	0.6
Oklahoma tornado outbreak 2010	OK	2010	10/5/2010	13/5/2010	0.6
Memorial Day tornado outbreak 2019	OH	2019	17/5/2019	31/5/2019	0.6
Nebraska flood 2019	NE	2019	13/3/2019	19/3/2019	0.6
Kansas tornado outbreak 2012	KS	2012	13/4/2012	16/4/2012	0.6
Oklahoma tornado outbreak sequence 2003	OK	2003	30/4/2003	11/5/2003	0.6
The 1998 Elba Flood	AL	1998	1/3/1998	31/3/1998	0.6
California flood 1995	CA	1995	1/1/1995	31/1/1995	0.6
Michigan windstorm 2017	MI	2017	7/3/2017	9/3/2017	0.6
Florida tornado 1998	FL	1998	21/2/1998	24/2/1998	0.5
Pennsylvania flood 2004	PA	2004	27/9/2004	30/9/2004	0.5
Idaho wildfire 2012	ID	2012	1/6/2012	30/6/2012	0.5
Colorado thunderstorm 2009	CO	2009	1/7/2009	31/7/2009	0.5
Arkansas tornado 1996	AR	1996	1/4/1996	30/4/1996	0.5
California storms and flooding 2006	CA	2006	1/1/2006	31/1/2006	0.5
California severe rainstorms 2005	CA	2005	1/12/2005	31/12/2005	0.5
Oklahoma hail 2012	OK	2012	28/5/2012	31/5/2012	0.5
Maine ice storm of 1998	ME	1998	4/1/1998	10/1/1998	0.5
Ohio flood 2006	OH	2006	27/7/2006	31/7/2006	0.5

# Chapter 6

## Conclusion and Future Works

### 6.1 Conclusion Remarks

This dissertation explores the role that actuarial science can play in addressing the multifaceted challenges posed by climate change and enhancing sustainability efforts within the financial and insurance industries. Through the integration of advanced computational models with actuarial practices, this thesis demonstrates how robust risk management strategies can be developed to cope with the growing climate-related challenges. Additionally, it highlights the potential of policy tools and climate-resilient assets from the financial market to enhance the sustainability of insurers and other financial institutions.

In Chapter 2, I demonstrate how a behavior-based machine learning approach can significantly improve the basis risk associated with area-yield insurance products, presenting a promising alternative to traditional individual yield insurance in the agricultural sector. This method not only mitigates moral hazard but also aligns with sustainable agricultural practices by supporting more accurate and fair insurance payouts.

Chapter 3 introduces a geo-hierarchical deep learning framework for flood risk pricing. This model has shown its potential to integrate climate and physical data into actuarial models at a lower cost and with greater robustness than traditional physical risk models, thereby enhancing the financial sustainability of flood insurance.

In Chapter 4, the focus shifted to the protection gap in climate risk insurance. The proposed self-financing tax redistribution framework offers a novel approach to mitigate this gap, highlighting the role of public policy in supporting private insurance solutions and fostering a more resilient societal structure against climate risks.

Finally, Chapter 5 illustrates how a corporation's ESG devotion can improve its climate sustainability by identifying valuation benefits during unfavorable climate conditions. The findings from natural experiments using hurricane strikes confirmed the greater climate resilience of green assets, which reduces insurers' insolvency risk and promotes a more sustainable investment strategy.

This thesis highlights the critical role of actuarial science in tackling climate change and advancing sustainability efforts. By incorporating advanced modeling techniques and emphasizing the need for climate-resilient assets, this research provides actionable insights for the insurance industry and policymakers. The proposed frameworks and models offer a pathway for developing effective risk management strategies that not only protect economic interests but also contribute to environmental and social resilience.

Looking ahead, climate change should be a focal point of future studies. The complexity and variability of climate systems necessitate continuous refinement of models and approaches to better predict and mitigate climate impacts. Actuaries and the insurance industry should consider the evolving nature of climate risks and adapt their practices accordingly. Further research should explore new data sources, innovative modeling techniques, and comprehensive policy interventions to enhance the resilience of financial systems against climate-related challenges.

## **6.2 Future Work Directions**

### **6.2.1 Alternative Dependence Structure for Area-yield Pooling**

In Chapter 2, the dependence structure among different producers is analyzed using a correlation matrix derived from historical yield records. An accurate dependence structure

is crucial for the effectiveness of the behavior-based risk pooling method proposed in the chapter. However, the reliance on extensive historical data presents a challenge, as sufficiently long time series data are not always available, which is a common scenario in practice. To address this, alternative dependence structure capturing methods, such as factor-based pooling, should be explored to enhance the robustness of the model, especially in situations where the credit history of producers is limited or inaccessible.

### **6.2.2 GHDL Model Application to Emerging Market**

The GHDL model introduced in Chapter 3 is designed to operate without requiring high-resolution or hard-to-access data, ensuring its applicability to emerging markets where data and resources may be limited. In the future, I plan to evaluate the performance of the GHDL-generated flood risk factors within the specific context of developing markets, such as in the creation of index insurance, to further validate and strengthen the model's predictive capabilities.

### **6.2.3 General Externality in Tax Redistribution Modeling**

In Chapter 4, I explore how uncovered climate losses in high-risk regions spill over into moderate-risk regions, using public disaster relief to model a conditional wealth transfer scheme. This scheme activates when climate losses in high-risk areas exceed a certain threshold, resulting in a wealth transfer from moderate to high-risk regions. While this represents a negative externality for moderate-risk regions and a positive one for high-risk regions, it is a limited perspective on externalities. For future work, it is necessary to analyze the effectiveness and optimization of the tax redistribution scheme in addressing the climate risk protection gap under broader types of externalities, including scenarios where both high-risk and moderate-risk regions experience negative externalities.

## 6.2.4 Optimal Public Sector Participation Rate

In Chapter 4, I discussed how, when the catastrophe probability  $r$  is exogenous, planners can determine their level of participation by selecting the required solvency capital level  $\alpha$ . Identifying a socially optimal  $\alpha$  that maximizes the overall welfare of both residents and insurers is a critical consideration. I will explore this topic further in future work.

## 6.2.5 Releasing Assumptions in Insurers' Financing and Investing

The propositions and theorems presented in Chapter 4 are based on two key assumptions: (1) insurers are prohibited from financing through credit, and (2) insurers invest their reserves exclusively in risk-free assets. These constraints are applied to simplify the discussion. However, the financing structure and asset portfolio of insurers can significantly influence their decisions regarding insurance supply. As part of future work, I plan to relax these assumptions to explore how they impact the findings of this chapter.

## 6.2.6 General Class of Green Assets' Climate Resilience

In Chapter 5, the tests for climate resilience were conducted on equity assets. While equities provide a good indication of a firm's resilience, it is important to assess the climate resilience of corporate bonds, which have a higher priority in payment hierarchies during a firm's financial distress caused by severe climate shocks. Moreover, corporate bonds typically constitute a larger portion of insurers' reserve portfolios compared to equities. Extending the findings of Chapter 5 to include a broader range of asset classes would enhance my understanding of the climate resilience properties of green assets.

## 6.2.7 Green Assets Climate Resilience Channel Testing

In the robustness analysis of Chapter 5, I explored the underlying mechanisms behind the enhanced resilience of green assets in the face of hurricane strikes. However, comprehensive investigations are necessary to thoroughly review the channels through which green assets

demonstrate climate resilience, as highlighted in this study. Future research should aim to delve deeper into these mechanisms to provide a more complete understanding of the resilience attributes of green assets against climate-related adversities.

# References

- Abbot, J. and Marohasy, J. (2014). Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using Artificial Neural Networks. *Atmospheric Research*, 138:166–178.
- Ackermann, F. (2024). Managing grand challenges: Extending the scope of problem structuring methods and behavioural operational research. *European Journal of Operational Research*.
- Aczel, M. (2022). Confronting climate injustice. *Science*, 376(6590):253–253.
- Addoum, J. M., Ng, D. T., and Ortiz-Bobea, A. (2020). Temperature shocks and establishment sales. *The Review of Financial Studies*, 33(3):1331–1366.
- Albuquerque, R., Koskinen, Y., and Zhang, C. (2019). Corporate social responsibility and firm risk: Theory and empirical evidence. *Management Science*, 65(10):4451–4469.
- Alok, S., Kumar, N., and Wermers, R. (2020). Do fund managers misestimate climatic disaster risk. *The Review of Financial Studies*, 33(3):1146–1183.
- Andor, M. A., Osberghaus, D., and Simora, M. (2020). Natural disasters and governmental aid: Is there a charity hazard? *Ecological Economics*, 169:106534.
- Annan, F., Tack, J., Harri, A., and Coble, K. (2014). Spatial pattern of yield distributions: Implications for crop insurance. *American Journal of Agricultural Economics*, 96(1):253–268.
- Arias, P., Bellouin, N., Coppola, E., Jones, R., Krinner, G., Marotzke, J., Naik, V., Palmer, M., Plattner, G.-K., Rogelj, J., Rojas, M., Sillmann, J., Storelvmo, T., Thorne, P., Trewin, B., Achuta Rao, K., Adhikary, B., Allan, R., Armour, K., Bala, G., Barimalala, R., Berger, S., Canadell, J., Cassou, C., Cherchi, A., Collins, W., Collins, W., Connors, S., Corti, S.,

- Cruz, F., Dentener, F., Dereczynski, C., Di Luca, A., Diongue Niang, A., Doblus-Reyes, F., Dosio, A., Douville, H., Engelbrecht, F., Eyring, V., Fischer, E., Forster, P., Fox-Kemper, B., Fuglestedt, J., Fyfe, J., Gillett, N., Goldfarb, L., Gorodetskaya, I., Gutierrez, J., Hamdi, R., Hawkins, E., Hewitt, H., Hope, P., Islam, A., Jones, C., Kaufman, D., Kopp, R., Kosaka, Y., Kossin, J., Krakovska, S., Lee, J.-Y., Li, J., Mauritsen, T., Maycock, T., Meinshausen, M., Min, S.-K., Monteiro, P., Ngo-Duc, T., Otto, F., Pinto, I., Pirani, A., Raghavan, K., Ranasinghe, R., Ruane, A., Ruiz, L., Sallée, J.-B., Samset, B., Sathyendranath, S., Seneviratne, S., Sörensson, A., Szopa, S., Takayabu, I., Tréguier, A.-M., van den Hurk, B., Vautard, R., von Schuckmann, K., Zaehle, S., Zhang, X., and Zickfeld, K. (2021). Technical summary. In [Masson-Delmotte et al. \(2021\)](#), pages 33–144.
- Balaguru, K., Xu, W., Chang, C.-C., Leung, L. R., Judi, D. R., Hagos, S. M., Wehner, M. F., Kossin, J. P., and Ting, M. (2023). Increased US coastal hurricane risk under climate change. *Science advances*, 9(14):eadf0259.
- Baldauf, M., Garlappi, L., and Yannelis, C. (2020). Does climate change affect real estate prices? Only if you believe in it. *The Review of Financial Studies*, 33(3):1256–1295.
- Bansal, R., Ochoa, M., and Kiku, D. (2016). Climate change and growth risks. *National Bureau of Economic Research*.
- Barberis, N. and Huang, M. (2008). Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review*, 98(5):2066–2100.
- Barberis, N., Huang, M., and Santos, T. (2001). Prospect theory and asset prices. *The quarterly journal of economics*, 116(1):1–53.
- Barko, P., McMichael, M., Swanson, K., and Williams, D. (2017). The gastrointestinal microbiome: A review. *Journal of Veterinary Internal Medicine*, 32(1):9–25.
- Barnett, B. J., Black, J. R., Hu, Y., and Skees, J. R. (2005). Is area yield insurance competitive with farm yield insurance? *Journal of Agricultural and Resource Economics*, 30(2):285–301.
- Barseghyan, L., Molinari, F., O’Donoghue, T., and Teitelbaum, J. C. (2013). The nature of risk preferences: Evidence from insurance choices. *American economic review*, 103(6):2499–2529.
- Benevolenza, M. A. and DeRigne, L. (2019). The impact of climate change and natural

- disasters on vulnerable populations: A systematic review of literature. *Journal of Human Behavior in the Social Environment*, 29(2):266–281.
- Beyer, K., Goldstein, J., Ramakrishnan, R., and Shaft, U. (1999). When is “nearest neighbor” meaningful? *Lecture Notes in Computer Science*, pages 217–235.
- Bolton, P. and Kacperczyk, M. (2021). Do investors care about carbon risk? *Journal of Financial Economics*, 142(2):517–549.
- Boonen, T. J., Chong, W. F., and Ghossoub, M. (2024). Pareto-efficient risk sharing in centralized insurance markets with application to flood risk. *Journal of Risk and Insurance*.
- Born, P., Cole, C. R., and Nyce, C. (2021). Citizens and the Florida residential property market: How to return to an insurer of last resort. *Journal of Insurance Regulation*.
- Botzen, W., Aerts, J., and van den Bergh, J. (2009). Willingness of homeowners to mitigate climate risk through insurance. *Ecological Economics*, 68(8–9):2265–2277.
- Boudreault, M., Grenier, P., Pigeon, M., Potvin, J.-M., and Turcotte, R. (2019). Pricing flood insurance with a hierarchical physics-based model. *North American Actuarial Journal*, 24(2):251–274.
- Boudreault, M. and Ojeda, A. (2022). Ratemaking territories and adverse selection for flood insurance. *Insurance: Mathematics and Economics*, 107:349–360.
- Bourgeon, J.-M. and Chambers, R. G. (2003). Optimal area-yield crop insurance reconsidered. *American Journal of Agricultural Economics*, 85(3):590–604.
- Bouwer, L. M. (2013). Projections of future extreme weather losses under changes in climate and exposure. *Risk Analysis*, 33(5):915–930.
- Browne, M. J., Dehring, C. A., Eckles, D. L., and Lastrapes, W. D. (2018). Does national flood insurance program participation induce housing development? *Journal of Risk and Insurance*, 86(4):835–859.
- Browne, M. J. and Hoyt, R. E. (2000). The demand for flood insurance: Empirical evidence. *Journal of Risk and Uncertainty*, 20(3):291–306.
- Bruggeman, V., Faure, M. G., and Fiore, K. (2010). The government as reinsurer of catastrophe risks? *The Geneva Papers on Risk and Insurance - Issues and Practice*, 35(3):369–390.
- Buchanan, J. M. (1975). *The Samaritan’s Dilemma*, page 71–86. Russell Sage Foundation.
- Cagle, J. A. and Harrington, S. E. (1995). Insurance supply with capacity constraints and

- endogenous insolvency risk. *Journal of Risk and Uncertainty*, 11(3):219–232.
- Cai, J., De Janvry, A., and Sadoulet, E. (2020). Subsidy policies and insurance demand. *American Economic Review*, 110(8):2422–53.
- Carter, M. R., Galarza, F., and Boucher, S. (2007). Underwriting area-based yield insurance to crowd-in credit supply demand. *Saving and Development*, 31(3):335–362.
- Casaburi, L. and Willis, J. (2018). Time versus state in insurance: Experimental evidence from contract farming in Kenya. *American Economic Review*, 108(12):3778–3813.
- Chambers, R. G. and Quiggin, J. (2002). Optimal producer behavior in the presence of area-yield crop insurance. *American Journal of Agricultural Economics*, 84(2):320–334.
- Chantararat, S., Mude, A. G., Barrett, C. B., and Carter, M. R. (2013). Designing index-based livestock insurance for managing asset risk in northern Kenya. *Journal of Risk and Insurance*, 80(1):205–237.
- Chantararat, S., Mude, A. G., Barrett, C. B., and Turvey, C. G. (2017). Welfare impacts of index insurance in the presence of a poverty trap. *World Development*, 94:119–138.
- Charpentier, A. (2007). Insurability of climate risks. *The Geneva Papers on Risk and Insurance - Issues and Practice*, 33(1):91–109.
- Charpentier, A. and Le Maux, B. (2014). Natural catastrophe insurance: How should the government intervene? *Journal of Public Economics*, 115:1–17.
- Chava, S. (2014). Environmental externalities and cost of capital. *Management Science*, 60(9):2223–2247.
- Chen, A. S., Djordjević, S., Leandro, J., and Savić, D. A. (2010). An analysis of the combined consequences of pluvial and fluvial flooding. *Water Science and Technology*, 62(7):1491–1498.
- Chen, D., Rojas, M., Samset, B., Cobb, K., Diongue Niang, A., Edwards, P., Emori, S., Faria, S., Hawkins, E., Hope, P., Huybrechts, P., Meinshausen, M., Mustafa, S., Plattner, G.-K., and Tréguier, A.-M. (2021). Framing, Context, and Methods. In [Masson-Delmotte et al. \(2021\)](#), page 147–286.
- Chen, T., Dong, H., and Lin, C. (2020). Institutional shareholders and corporate social responsibility. *Journal of Financial Economics*, 135(2):483–504.
- Chen, Z., Lu, Y., Zhang, J., and Zhu, W. (2023a). Managing weather risk with a neural

- network-based index insurance. *Management Science*, 0(0):1–22.
- Chen, Z., Lu, Y., Zhang, J., and Zhu, W. (2023b). Managing weather risk with a neural network-based index insurance. *Management Science*, *forthcoming*.
- Choi, D., Gao, Z., and Jiang, W. (2020). Attention to global warming. *The Review of Financial Studies*, 33(3):1112–1145.
- Choubin, B., Zehtabian, G., Azareh, A., Rafiei-Sardooi, E., Sajedi-Hosseini, F., and Kişi, O. (2018). Precipitation forecasting using classification and regression trees (CART) model: A comparative study of different approaches. *Environmental Earth Sciences*, 77(8).
- Clarke, D. J. (2016). A theory of rational demand for index insurance. *American Economic Journal: Microeconomics*, 8(1):283–306.
- Cole, S., Stein, D., and Tobacman, J. (2014). Dynamics of demand for index insurance: Evidence from a long-run field experiment. *American Economic Review*, 104(5):284–90.
- Collier, B. L., Schwartz, D., Kunreuther, H. C., and Michel-Kerjan, E. O. (2022). Insuring large stakes: A normative and descriptive analysis of households’ flood insurance coverage. *Journal of Risk and Insurance*, 89(2):273–310.
- Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichet, T., Friedlingstein, P., Gao, X., Gutowski, W., et al. (2013). Long-term climate change: Projections, commitments and irreversibility. In IPCC, editor, *Climate Change 2013: The Physical Science Basis. IPCC Working Group I Contribution to AR5*, chapter 12. Cambridge University Press, Cambridge.
- Conell-Price, L., Kousky, C., and Kunreuther, H. (2022). Encouraging resiliency through autoenrollment in supplemental flood insurance coverage. *Journal of Risk and Insurance*, 89(4):1109–1137.
- Cong, L. W., Liang, T., and Zhang, X. (2019). Textual factors: A scalable, interpretable, and data-driven approach to analyzing unstructured information. *Interpretable, and Data-driven Approach to Analyzing Unstructured Information (September 1, 2019)*.
- Cuñat, V., Gine, M., and Guadalupe, M. (2012). The vote is cast: The effect of corporate governance on shareholder value. *The Journal of Finance*, 67(5):1943–1977.
- Cummins, J. D. (2006). Should the government provide insurance for catastrophes? *Review*, 88(4).

- Cummins, J. D., Lalonde, D., and Phillips, R. D. (2004). The basis risk of catastrophic-loss index securities. *Journal of Financial Economics*, 71(1):77–111.
- Cummins, J. D. and Weiss, M. A. (2000). Analyzing firm performance in the insurance industry using frontier efficiency and productivity methods. *Handbook of Insurance*, page 767–829.
- Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41(7):909–996.
- Dehghani, M., Saghafian, B., Rivaz, F., and Khodadadi, A. (2017). Evaluation of dynamic regression and artificial neural networks models for real-time hydrological drought forecasting. *Arabian Journal of Geosciences*, 10(12).
- Deng, X., Barnett, B. J., and Vedenov, D. V. (2007). Is there a viable market for area-based crop insurance? *American Journal of Agricultural Economics*, 89(2):508–519.
- Deng, X., Kang, J.-k., and Low, B. S. (2013). Corporate social responsibility and stakeholder value maximization: Evidence from mergers. *Journal of Financial Economics*, 110(1):87–109.
- Dessaint, O. and Matray, A. (2017). Do managers overreact to salient risks? Evidence from hurricane strikes. *Journal of Financial Economics*, 126(1):97–121.
- Dietz, S., Gollier, C., and Kessler, L. (2018). The climate beta. *Journal of Environmental Economics and Management*, 87:258–274.
- Dimson, E., Karakaş, O., and Li, X. (2015). Active ownership. *Review of Financial Studies*, 28(12):3225–3268.
- Eby, M. (2023). NOAA’s 1-in-100 year flooding can now be expected every 8 years. *First Street Foundation*.
- Edmans, A. (2011). Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial Economics*, 101(3):621–640.
- EIOPA (2023). European Insurance and Occupational Pensions Authority: Solvency II. [https://www.eiopa.europa.eu/browse/regulation-and-policy/solvency-ii\\_en](https://www.eiopa.europa.eu/browse/regulation-and-policy/solvency-ii_en). Accessed: 2023-11-03.
- El Ghouli, S., Guedhami, O., Kwok, C. C., and Mishra, D. R. (2011). Does corporate social responsibility affect the cost of capital? *Journal of Banking & Finance*, 35(9):2388–2406.

- El Ghouli, S., Guedhami, O., Wang, H., and Kwok, C. C. (2016). Family control and corporate social responsibility. *Journal of Banking & Finance*, 73:131–146.
- Elabed, G., Bellemare, M. F., Carter, M. R., and Guirking, C. (2013). Managing basis risk with multiscale index insurance. *Agricultural Economics*, 44(4-5):419–431.
- Elabed, G. and Carter, M. R. (2015). Compound-risk aversion, ambiguity and the willingness to pay for microinsurance. *Journal of Economic Behavior & Organization*, 118:150–166.
- Engle, R. F., Giglio, S., Kelly, B., Lee, H., and Stroebe, J. (2020). Hedging climate change news. *The Review of Financial Studies*, 33(3):1184–1216.
- Er Kara, M., Ghadge, A., and Bititci, U. S. (2020). Modelling the impact of climate change risk on supply chain performance. *International Journal of Production Research*, 59(24):7317–7335.
- Flammer, C. (2015). Does corporate social responsibility lead to superior financial performance? A regression discontinuity approach. *Management Science*, 61(11):2549–2568.
- Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., et al. (2014). Evaluation of climate models. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 741–866. Cambridge University Press.
- Flavelle, C. (2022). Hurricane Ian’s toll is severe. Lack of insurance will make it worse. *The New York Times*.
- Fox-Kemper, B., Hewitt, H., Xiao, C., Aðalgeirsdóttir, G., Drijfhout, S., Edwards, T., Gollidge, N., Hemer, M., Kopp, R., Krinner, G., Mix, A., Notz, D., Nowicki, S., Nurhati, I., Ruiz, L., Sallée, J.-B., Slangen, A., and Yu, Y. (2021). Ocean, Cryosphere and Sea Level Change. In [Masson-Delmotte et al. \(2021\)](#), pages 1211–1362.
- Fulga, C. (2016). Portfolio optimization under loss aversion. *European journal of operational research*, 251(1):310–322.
- Gall, M. (2023). Why insurance companies are pulling out of California and Florida, and how to fix some of the underlying problems.
- Ge, S. (2021). How do financial constraints affect product pricing? Evidence from weather and life insurance premiums. *The Journal of Finance*, 77(1):449–503.

- Ghadge, A., Wurtmann, H., and Seuring, S. (2019). Managing climate change risks in global supply chains: A review and research agenda. *International Journal of Production Research*, 58(1):44–64.
- GIIF (2024). What are the different types of crop index insurance? Accessed: 2024-08-01.
- Gillan, S. (2010). Firms’ environmental, social and governance (ESG) choices, performance and managerial motivation. *Unpublished Working Paper*.
- Gong, X., Hennessy, D. A., and Feng, H. (2023). Systemic risk, relative subsidy rates, and area yield insurance choice. *American Journal of Agricultural Economics*, 105(3):888–913.
- Goodwin, B. K. (1993). An empirical analysis of the demand for multiple peril crop insurance. *American Journal of Agricultural Economics*, 75(2):425–434.
- Goss, A. and Roberts, G. S. (2011). The impact of corporate social responsibility on the cost of bank loans. *Journal of Banking & Finance*, 35(7):1794–1810.
- Goswami, A., Borasi, P., and Kumar, V. (2020). Crop insurance market by coverage and distribution channel : Global opportunity analysis and industry forecast, 2020–2027. Technical report, Allied Market Research.
- Guo, Z., Leitão, J. P., Simões, N. E., and Moosavi, V. (2020). Data-driven flood emulation: Speeding up urban flood predictions by deep convolutional neural networks. *Journal of Flood Risk Management*, 14(1).
- Hao, M., Macdonald, A. S., Tapadar, P., and Thomas, R. G. (2018). Insurance loss coverage and demand elasticities. *Insurance: Mathematics and Economics*, 79(1):15–25.
- Härdle, W., Kerkycharian, G., Picard, D., and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications*, volume 129 of *Lecture Notes in Statistics*. Springer.
- Harri, A., Coble, K. H., Ker, A. P., and Goodwin, B. J. (2011a). Relaxing heteroscedasticity assumptions in area-yield crop insurance rating. *American Journal of Agricultural Economics*, 93(3):707–717.
- Harri, A., Coble, K. H., Ker, A. P., and Goodwin, B. J. (2011b). Relaxing heteroscedasticity assumptions in area-yield crop insurance rating. *American Journal of Agricultural Economics*, 93(3):707–717.
- Harris, R. D. and Mazibas, M. (2022). Portfolio optimization with behavioural preferences and investor memory. *European Journal of Operational Research*, 296(1):368–387.

- Hawkins, E. and Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, 90(8):1095–1108.
- Hazell, P., Jaeger, A., and Hausberger, R. (2021). Innovations and emerging trends in agricultural insurance: An update. <https://www.giz.de>. Accessed: 2024-06-07.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- Helldén, D., Andersson, C., Nilsson, M., Ebi, K. L., Friberg, P., and Alfvén, T. (2021). Climate change and child health: a scoping review and an expanded conceptual framework. *The Lancet Planetary Health*, 5(3):e164–e175.
- Henderson, M. (2023). 2023 Louisiana survey shows homeowners face insurance challenges.
- Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., and Kanae, S. (2013). Global flood risk under climate change. *Nature Climate Change*, 3(9):816–821.
- Hoepner, A. G., Oikonomou, I., Sautner, Z., Starks, L. T., and Zhou, X. Y. (2023). ESG shareholder engagement and downside risk. *Review of Finance*, 28(2):483–510.
- Holland, G. and Bruyère, C. L. (2013). Recent intense hurricane response to global climate change. *Climate Dynamics*, 42(3–4):617–627.
- Hong, H., Li, F. W., and Xu, J. (2019). Climate risks and market efficiency. *Journal of Econometrics*, 208(1):265–281.
- Hossain, M. S., Alam, G. M., Fahad, S., Sarker, T., Moniruzzaman, M., and Rabbany, M. G. (2022). Smallholder farmers’ willingness to pay for flood insurance as climate change adaptation strategy in northern Bangladesh. *Journal of Cleaner Production*, 338:130584.
- Hu, Z. (2022). Social interactions and households’ flood insurance decisions. *Journal of Financial Economics*, 144(2):414–432.
- Hudak, M. (2022). Floridians running out of options for home insurance.
- Hwang, I., Xu, S., and In, F. (2018). Naive versus optimal diversification: Tail risk and performance. *European Journal of Operational Research*, 265(1):372–388.
- Ilhan, E., Krueger, P., Sautner, Z., and Starks, L. T. (2019). Institutional investors’ views and preferences on climate risk disclosure. *SSRN Electronic Journal*.

- IPCC (2021). Summary for Policymakers. In [Masson-Delmotte et al. \(2021\)](#), pages 3–32.
- Isidore, C. (2023). Florida’s homeowner insurance rates are four times the national average. That’s not getting better anytime soon.
- Isidore, C. and Nilsen, E. (2023). Why it’s becoming harder and more expensive to get homeowners insurance.
- Iturrioz, R. (2009). Agricultural insurance. *Primer Series on Insurance. The World Bank.*, No. E20-77:1–35.
- Jensen, N. D., Barrett, C. B., and Mude, A. G. (2016). Index insurance quality and basis risk: Evidence from northern Kenya. *American Journal of Agricultural Economics*, 98(5):1450–1469.
- Jiao, Y. (2010). Stakeholder welfare and firm value. *Journal of Banking & Finance*, 34(10):2549–2561.
- Jiraporn, P., Jiraporn, N., Boeprasert, A., and Chang, K. (2014). Does corporate social responsibility (CSR) improve credit ratings? Evidence from geographic identification. *Financial Management*, 43(3):505–531.
- Jongman, B., Hochrainer-Stigler, S., Feyen, L., Aerts, J. C., Mechler, R., Botzen, W. W., Bouwer, L. M., Pflug, G., Rojas, R., and Ward, P. J. (2014). Increasing stress on disaster-risk finance due to large floods. *Nature Climate Change*, 4(4):264–268.
- Ker, A. P., Tolhurst, T. N., and Liu, Y. (2015). Bayesian estimation of possibly similar yield densities: Implications for rating crop insurance contracts. *American Journal of Agricultural Economics*, 98(2):360–382.
- Kousky, C. and Kunreuther, H. (2018). Risk management roles of the public and private sector. *Risk Management and Insurance Review*, 21(1):181–204.
- Kousky, C., Lingle, B., and Shabman, L. (2017). The pricing of flood insurance. *Journal of Extreme Events*, 04(02):1750001.
- Kousky, C. and Michel-Kerjan, E. (2017). Examining flood insurance claims in the United States: Six key findings. *Journal of Risk and Insurance*, 84(3):819–850.
- Kreibich, H., Van Loon, A. F., Schröter, K., Ward, P. J., Mazzoleni, M., Sairam, N., Abeshu, G. W., Agafonova, S., AghaKouchak, A., Aksoy, H., et al. (2022). The challenge of unprecedented floods and droughts in risk management. *Nature*, 608(7921):80–86.

- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.
- Krueger, P., Sautner, Z., and Starks, L. T. (2020). The importance of climate risks for institutional investors. *The Review of Financial Studies*, 33(3):1067–1111.
- Kumar, S., Tiwari, M. K., Chatterjee, C., and Mishra, A. (2015). Reservoir inflow forecasting using ensemble models based on neural networks, wavelet analysis and Bootstrap Method. *Water Resources Management*, 29(13):4863–4883.
- Kunreuther, H. (2015). The role of insurance in reducing losses from extreme events: The need for public–private partnerships. *The Geneva Papers on Risk and Insurance - Issues and Practice*, 40(4):741–762.
- Kusumaningrum, D., Anisa, R., Sutomo, V. A., and Tan, K. S. (2021). Alternative area yield index based crop insurance policies in indonesia. In *Mathematical and Statistical Methods for Actuarial Sciences and Finance: eMAF2020*, pages 285–290. Springer.
- Lechowska, E. (2018). What determines flood risk perception? A review of factors of flood risk perception and relations between its basic elements. *Natural Hazards*, 94(3):1341–1366.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- Lee, J.-Y., Marotzke, J., Bala, G., Cao, L., Corti, S., Dunne, J., Engelbrecht, F., Fischer, E., Fyfe, J., Jones, C., Maycock, A., Mutemi, J., Ndiaye, O., Panickal, S., and Zhou, T. (2021). Future Global Climate: Scenario-Based Projections and Near-Term Information. In [Masson-Delmotte et al. \(2021\)](#), pages 553–672.
- Leefeldt, E. (2024). Why is homeowners insurance in Florida such a disaster?
- Levy, B. S. and Patz, J. A. (2015). Climate change, human rights, and social justice. *Annals of Global Health*, 81(3):310.
- Li, L., Xu, H., Chen, X., and Simonovic, S. P. (2009). Streamflow forecast and reservoir operation performance assessment under climate change. *Water Resources Management*, 24(1):83–104.
- Li, Z., Liu, F., Yang, W., Peng, S., and Zhou, J. (2022). A survey of Convolutional Neural Networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks*

- and Learning Systems*, 33(12):6999–7019.
- Lin, N., Kopp, R. E., Horton, B. P., and Donnelly, J. P. (2016). Hurricane Sandy’s flood frequency increasing from year 1800 to 2100. *Proceedings of the National Academy of Sciences*, 113(43):12071–12075.
- Lins, K. V., Servaes, H., and Tamayo, A. (2017). Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. *The Journal of Finance*, 72(4):1785–1824.
- Liu, Y. and Ker, A. P. (2021). Simultaneous borrowing of information across space and time for pricing insurance contracts: An application to rating crop insurance policies. *Journal of Risk and Insurance*, 88(1):231–257.
- Lowder, S. K., Skoet, J., and Raney, T. (2016). The number, size, and distribution of farms, smallholder farms, and family farms worldwide. *World development*, 87:16–29.
- Luo, H. A. and Balvers, R. J. (2017). Social screens and systematic investor boycott risk. *Journal of Financial and Quantitative Analysis*, 52(1):365–399.
- Mahul, O. (1999). Optimum area yield crop insurance. *American Journal of Agricultural Economics*, 81(1):75–82.
- Mahul, O. and Stutley, C. J. (2010). *Government support to agricultural insurance: Challenges and options for developing countries*. World Bank Publications.
- Mann, M. E. and Emanuel, K. A. (2006). Atlantic hurricane trends linked to climate change. *Eos, Transactions American Geophysical Union*, 87(24):233–241.
- Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., editors (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Michel-Kerjan, E. and Kunreuther, H. (2011). Redesigning flood insurance. *Science*, 333(6041):408–409.
- Michel-Kerjan, E. O. and Kousky, C. (2010). Come rain or shine: Evidence on flood insurance purchases in Florida. *Journal of Risk and Insurance*, 77(2):369–397.

- Mintz, Y., Aswani, A., Kaminsky, P., Flowers, E., and Fukuoka, Y. (2023). Behavioral analytics for myopic agents. *European Journal of Operational Research*, 310(2):793–811.
- Miranda, M. J. (1991). Area-yield crop insurance reconsidered. *American Journal of Agricultural Economics*, 73(2):233–242.
- Miranda, M. J. and Glauber, J. (1997). Systemic risk, reinsurance and the failure of crop insurance market. *American Journal of Agriculture Economics*, 79(1):206–215.
- Morris, C. and Botros, A. (2023). 4 more insurers are leaving California after Allstate and State Farm flee wildfire risk.
- Mosavi, A., Ozturk, P., and Chau, K.-w. (2018). Flood prediction using machine learning models: Literature review. *Water*, 10(11):1536.
- Mukherjee, S., Mishra, A., and Trenberth, K. E. (2018). Climate change and drought: a perspective on drought indices. *Current climate change reports*, 4:145–163.
- Nagurney, A., Hassani, D., Nivievskiy, O., and Martyshev, P. (2024). Multicommodity international agricultural trade network equilibrium: Competition for limited production and transportation capacity under disaster scenarios with implications for food security. *European Journal of Operational Research*, 314(3):1127–1142.
- Netusil, N. R., Kousky, C., Neupane, S., Daniel, W., and Kunreuther, H. (2021). The willingness to pay for flood insurance. *Land Economics*, 97(1):17–38.
- NOAA (2023). *NOAA ATLAS 15: Update to the National Precipitation Frequency Standard*. [https://www.weather.gov/media/owp/hdsc\\_documents/NOAA\\_Atlas\\_15\\_Flyer.pdf](https://www.weather.gov/media/owp/hdsc_documents/NOAA_Atlas_15_Flyer.pdf). Accessed: 2024-01-06.
- Pankratz, N. M. and Schiller, C. M. (2022). Climate change and adaptation in global supply-chain networks. *Finance and Economics Discussion Series*, 2022(056).
- Pástor, v., Stambaugh, R. F., and Taylor, L. A. (2021). Sustainable investing in equilibrium. *Journal of Financial Economics*, 142(2):550–571.
- Pedersen, L. H., Fitzgibbons, S., and Pomorski, L. (2021). Responsible investing: The ESG-efficient frontier. *Journal of Financial Economics*, 142(2):572–597.
- Penning-Rowsell, E. C. (2015). Flood insurance in the uk: A critical perspective. *WIREs Water*, 2(6):601–608.
- Pielke Jr, R. A., Gratz, J., Landsea, C. W., Collins, D., Saunders, M. A., and Musulin, R.

- (2008). Normalized hurricane damage in the United States: 1900–2005. *Natural hazards review*, 9(1):29–42.
- Porth, L. and Tan, K. S. (2015). Agricultural insurance—more room to grow? *The Actuary*, April/May 2015 Issue.
- Porth, L., Tan, K. S., and Zhu, W. (2019). A relational data matching model for enhancing individual loss experience: An example from crop insurance. *North American Actuarial Journal*, 23(4):551–572.
- Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, 32(1/2):122.
- Quinn, N., Bates, P. D., Neal, J., Smith, A., Wing, O., Sampson, C., Smith, J., and Hefernan, J. (2019). The spatial dependence of flood hazard and risk in the United States. *Water Resources Research*, 55(3):1890–1911.
- Rahman, K. (2023). Florida insurance crisis explained: Why multiple insurers are leaving state.
- Ramaswami, B. and Roe, T. L. (2004). Aggregation in area-yield crop insurance: The linear additive model. *American Journal of Agricultural Economics*, 86(2):420–431.
- Raschky, P. and Weckhannemann, H. (2007). Charity hazard—a real hazard to natural disaster insurance? *Environmental Hazards*, 7(4):321–329.
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788.
- Rocque, R. J., Beaudoin, C., Ndjaboue, R., Cameron, L., Poirier-Bergeron, L., Poulin-Rheault, R.-A., Fallon, C., Tricco, A. C., and Witteman, H. O. (2021). Health effects of climate change: An overview of systematic reviews. *BMJ Open*, 11(6).
- Romanello, M., Di Napoli, C., Drummond, P., Green, C., Kennard, H., Lampard, P., Scamman, D., Arnell, N., Ayeb-Karlsson, S., Ford, L. B., and et al. (2022). The 2022 report of the lancet countdown on health and climate change: Health at the mercy of fossil fuels. *The Lancet*, 400(10363):1619–1654.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

- Rosch, S. (2021). Federal crop insurance: A primer. *Congressional Research Service*.
- Rowell, D. and Connelly, L. B. (2012). A history of the term “moral hazard”. *Journal of Risk and Insurance*, 79(4):1051–1075.
- Saintilan, N., Khan, N., Ashe, E., Kelleway, J., Rogers, K., Woodroffe, C. D., and Horton, B. P. (2020). Thresholds of mangrove survival under rapid sea level rise. *Science*, 368(6495):1118–1121.
- Sang, Y., Zhang, H., and Zuo, L. (2008). Least squares support vector machine classifiers using pcnns. *2008 IEEE Conference on Cybernetics and Intelligent Systems*.
- Santana, R. and Phillis, M. (2022). Lack of flood insurance leaves families with broken homes following Hurricane Ian. *Public Broadcasting Service*.
- Schlenker, W. and Taylor, C. A. (2021). Market expectations of a warming climate. *Journal of Financial Economics*, 142(2):627–640.
- Schubert, E., Koos, A., Emrich, T., Züfle, A., Schmid, K. A., and Zimek, A. (2015). A framework for clustering uncertain data. *Proc. VLDB Endow.*, 8(12):1976–1979.
- Seltzer, L., Starks, L. T., and Zhu, Q. (2020). Climate regulatory risks and corporate bonds. *SSRN Electronic Journal*.
- Seo, Y., Kim, S., Kisi, O., and Singh, V. P. (2015). Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *Journal of Hydrology*, 520:224–243.
- Sethanand, K., Chaiyawat, T., and Gowanit, C. (2023). Systematic process for crop insurance development: area-yield rice insurance with machine learning technology implementation in thailand. *Agricultural Finance Review*, 83(3):416–436.
- Shafaei, M. and Kisi, O. (2016). Predicting river daily flow using wavelet-artificial neural networks based on regression analyses in comparison with artificial neural networks and support vector machine models. *Neural Computing and Applications*, 28(S1):15–28.
- Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., Hegerl, G., Klein, S. A., Marvel, K. D., Rohling, E. J., et al. (2020). An assessment of Earth’s climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, 58(4):e2019RG000678.
- Shin, H.-H. and Stulz, R. M. (1998). Are internal capital markets efficient? *The Quarterly Journal of Economics*, 113(2):531–552.

- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Skees, J. R., Black, J. R., and Barnett, B. J. (1997). Designing and rating an area yield crop insurance contract. *American Journal of Agricultural Economics*, 79(2):430–438.
- Smith, V. H., Chouinard, H. H., and Baquet, A. E. (1994). Almost ideal area yield crop insurance contracts. *Agricultural and Resource Economics Review*, 23(1):75–83.
- Stein, J. C. (1997). Internal capital markets and the competition for corporate resources. *The Journal of Finance*, 52(1):111–133.
- Strauss, B. H., Orton, P. M., Bittermann, K., Buchanan, M. K., Gilford, D. M., Kopp, R. E., Kulp, S., Massey, C., Moel, H. d., and Vinogradov, S. (2021). Economic damages from Hurricane Sandy attributable to sea level rise caused by anthropogenic climate change. *Nature communications*, 12(1):2720.
- Surminski, S., Aerts, J. C., Botzen, W. J., Hudson, P., Mysiak, J., and Pérez-Blanco, C. D. (2015). Reflections on the current debate on how to link flood insurance and disaster risk reduction in the European Union. *Natural Hazards*, 79:1451–1479.
- Swiss Re (2021). Natural catastrophes in 2020: secondary perils in the spotlight, but don’t forget primary-peril risks. *Swiss Re sigma*. Accessed: 2024-08-01.
- Swiss Re (2022). Natural catastrophes in 2021: the floodgates are open. <https://www.swissre.com/institute/research/sigma-research/sigma-2022-01.html>. Accessed: 2023-11-03.
- Swiss Re (2024). Continued high losses from natural catastrophes in 2022.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9.
- Tack, J. B. and Ubilava, D. (2015). Climate and agricultural risk: Measuring the effect of ENSO on U.S. crop insurance. *Agricultural Economics*, 46(2):245–257.
- Tesselaar, M., Botzen, W. W., Robinson, P. J., Aerts, J. C., and Zhou, F. (2022). Charity hazard and the flood insurance protection gap: An EU scale assessment under climate change. *Ecological Economics*, 193:107289.
- The Guardian (2023). “Cascading impacts” warning as farmers becomes latest insurer to

quit Florida.

- Thistlethwaite, J., Henstra, D., Brown, C., and Scott, D. (2020). Barriers to insurance as a flood risk management tool: Evidence from a survey of property owners. *International Journal of Disaster Risk Science*, 11(3):263–273.
- Thomas, K., Hardy, R. D., Lazrus, H., Mendez, M., Orlove, B., Rivera-Collazo, I., Roberts, J. T., Rockman, M., Warner, B. P., and Winthrop, R. (2018). Explaining differential vulnerability to climate change: A social science review. *WIREs Climate Change*, 10(2).
- Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423.
- Tolan, C. and Devine, C. (2022). Lack of flood insurance in hard-hit Central Florida leaves families struggling after Hurricane Ian. *CNN*.
- UN (2023). Goal 13: Take urgent action to combat climate change and its impacts.
- Venturini, A. (2022). Climate change, risk factors and stock returns: A review of the literature. *International Review of Financial Analysis*, 79:101934.
- Vercammen, J. A. (2000). Constrained efficient contracts for area yield crop insurance. *American Journal of Agricultural Economics*, 82(4):856–864.
- Wang, H. H. and Zhang, H. (2003). On the possibility of a private crop insurance market: A spatial statistics approach. *The Journal of Risk and Insurance*, 70(1):111–124.
- Westerling, A. L. and Bryant, B. P. (2007). Climate change and wildfire in California. *Climatic Change*, 87(S1):231–249.
- WHO (2023). Climate change.
- Wing, O. E., Pinter, N., Bates, P. D., and Kousky, C. (2020). New insights into US flood vulnerability revealed from flood insurance big data. *Nature Communications*, 11(1).
- Woodard, J. D. (2016). Data science and management for large scale empirical applications in agricultural and applied economics research. *Applied Economic Perspectives and Policy*, 38(3):373–388.
- Woodard, J. D. and Garcia, P. (2008). Weather derivatives, spatial aggregation, and systemic risk: Implications for reinsurance hedging. *Journal of Agricultural and Resource Economics*, 33(1):34–51.

- World Bank (2023). World bank country and lending groups. <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>. Accessed: 2024-06-07.
- World Economic Forum (2023). Global risks report 2023. <https://www.weforum.org/publications/global-risks-report-2023/>. Accessed: 2024-05-20.
- Wu, C. L., Chau, K. W., and Li, Y. S. (2009). Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques. *Water Resources Research*, 45(8).
- Ye, T., Hu, W., Barnett, B. J., Wang, J., and Gao, Y. (2020). Area yield index insurance or farm yield crop insurance? Chinese perspectives on farmers' welfare and government subsidy effectiveness. *Journal of Agricultural Economics*, 71(1):144–164.
- Zhang, J., Li, Y., Xu, H., and Ding, Y. (2023). Can ESG ratings mitigate managerial myopia? Evidence from Chinese listed companies. *International Review of Financial Analysis*, 90:102878.
- Zhang, J.-S. and Xiao, X.-C. (2000). Predicting chaotic time series using Recurrent Neural Network. *Chinese Physics Letters*, 17(2):88–90.
- Zhu, W., Tan, K. S., and Porth, L. (2019). Agricultural insurance ratemaking: Development of a new premium principle. *North American Actuarial Journal*, 23(4):512–534.
- Zhu, W., Tan, K. S., Porth, L., and Wang, C.-W. (2018). Spatial dependence and aggregation in weather risk hedging: A Lévy subordinated hierarchical Archimedean copulas (LSHAC) approach. *ASTIN Bulletin: The Journal of the IAA*, 48(2):779–815.

# Glossary

**Adverse selection** A situation in which an insurer is unable to differentiate between high-risk and low-risk individuals, leading to a disproportionate number of high-risk policyholders and resulting in higher-than-expected losses [6](#)

**Basis risk** The risk that the payout from an [index insurance](#) policy does not match the actual loss experienced by the policyholder [8](#)

**Capacity** In the context of insurance, the maximum amount of risk an insurance company can underwrite, determined by the insurer's financial strength, regulatory requirements, risk appetite, capital and surplus, reinsurance, regulatory limits, risk appetite, and exposure management first [89](#)

**Cost of capital** The cost of funds used for financing a business, representing the return expected by investors for providing capital to the company [121](#)

**Credit risk** The risk of loss resulting from a borrower's failure to repay a loan or meet contractual obligations [121](#)

**Downside risk** The potential for loss in value of an investment or portfolio, representing the risk of the actual return being lower than the expected return [121](#)

**Expected shortfall** A risk measure that takes into account the size of losses exceeding the [Value-at-Risk](#) (VaR). The formula for ES at confidence level  $\alpha$  is:

$$ES_{\alpha}(X) = E[X \mid X \geq \text{VaR}_{\alpha}(X)]$$

where  $X$  is the loss random variable [27](#)

**Expected utility maximizer** An individual or entity that makes decisions to maximize their expected utility, a concept in economics and game theory where choices are made based on the expected outcomes [92](#)

**Indemnity insurance** A type of insurance in which the insurer compensates the insured for the actual loss suffered, up to the limit of the insurance policy [6](#)

**Index insurance** A type of insurance in which payouts are based on a predetermined index (such as rainfall level or temperature) rather than the actual loss incurred by the insured [6](#), [184](#)

**Mean squared error** A measure of the average of the squares of the errors or deviations, i.e., the difference between the estimator and what is estimated. The formula is given by:

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

where  $y_i$  is the actual value and  $\hat{y}_i$  is the estimated value [26](#)

**Moral hazard** The tendency of an insured party to take on more risk because they do not bear the full consequences of that risk [6](#)

**Non-pecuniary utility** The satisfaction or benefit derived from goods, services, or activities that are not measured in monetary terms, such as enjoyment, convenience, or personal fulfillment [123](#)

**Premium loading factor** The additional amount included in an insurance premium to cover expenses, contingencies, and profit margins [107](#)

**Price taker** An individual or company that must accept prevailing prices in a market, lacking the power to influence market prices [92](#)

**Required rate of return** The minimum return an investor expects to achieve by investing in a particular asset, considering its risk level [97](#)

**Risk averse** A term used to describe an investor or individual who prefers lower returns with known risks rather than higher returns with unknown risks [92](#)

**Risk neutral** A term used to describe an investor or individual who is indifferent to risk when making an investment decision, focusing solely on potential returns [92](#)

**Risk prudent** A term used to describe an individual or entity that takes careful and calculated risks, often incorporating a margin of safety in their decisions [107](#)

**Risk-free asset** An asset that is assumed to have no risk of financial loss, typically exemplified by government bonds or treasury bills, which are considered to have the lowest default risk [96](#)

**Risk-free rate of return** The theoretical return on an investment with zero risk, often represented by the yield on government bonds [96](#)

**Tail risk** The risk of rare events that have a significant impact on an asset. It refers to the extreme changes in asset prices or outcomes that occur at the tails of the probability distribution of returns [9](#)

**Underwriting loss** The financial loss an insurance company experiences when the claims and expenses exceed the premiums collected from policyholders [6](#)

**Value-at-Risk** A measure of the risk of loss for investments. It estimates how much the value of an asset will loss at the worst  $\alpha$  probability. The formula for VaR at confidence level  $\alpha$  is:

$$\text{VaR}_\alpha(X) = \inf\{x \in \mathbb{R} : P(X \leq x) \geq \alpha\}$$

where  $X$  is the loss random variable [26](#), [184](#)