ANATOMY OF FLOOD RISK AND FLOOD INSURANCE IN THE U.S.

by

Arkaprabha Bhattacharyya

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Lyles School of Civil Engineering

West Lafayette, Indiana

December 2023

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Makarand Hastak, Chair
Lyles School of Civil Engineering

Dr. Satish V. Ukkusuri
Lyles School of Civil Engineering

Dr. Holly H. Wang
School of Agricultural Economics

Dr. David Yu
Lyles School of Civil Engineering

Approved by:
Dr. Dulcy M. Abraham

"It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of light, it was the season of darkness, it was the spring of hope, it was the winter of despair."

— *Charles Dickens, A Tale of Two Cities*

Dedicated to my family, whose unwavering support has been instrumental in this unique journey of self-discovery.

ACKNOWLEDGEMENTS

This research journey, characterized by hope and despair, owes its realization to the invaluable contributions of many individuals whom I regard as my friends, philosophers, and guides. Foremost, I extend my gratitude to my advisor, Dr. Makarand Hastak, whose unwavering guidance and support have sustained me over the past five years. When I initially joined his lab, I was a novice in academic research, yet he recognized my potential and introduced me to the domain of disaster risk reduction. Throughout this period, his support for my research endeavors and patience with my research output have been monumental to my accomplishments at Purdue University. His mentorship has been exceptional, with a keen focus on my growth. Without his guidance, this research would not have materialized.

I am equally appreciative of the insights offered by my committee members. Given the interdisciplinary nature of this research, I am indebted to Dr. Holly Wang for her contributions, particularly in the realm of Economics. I am thankful to Dr. Satish Ukkusuri for both his feedback on this research and his illuminating graduate-level course on Data Science for Smart Cities, which expanded my horizons. Dr. David Yu's constructive reviews have also been invaluable.

I extend heartfelt gratitude to Dr. Doug Gotham, Director of the State Utility Forecasting Group at Purdue University, for his continuous support through my research assistant appointment. His mentorship over the past two years has been transformative, with his accessibility and care greatly appreciated. I also wish to acknowledge Dr. Bob Bowen, under whom I served as a teaching assistant for three year between 2019 to 2021. His recognition of my contributions and his empowerment of my efforts have taught me the importance of fostering productivity in others. Dr. Soojin Yoon has been instrumental in my transition from a graduate student to a graduate research assistant, and I am thankful for her guidance.

I express gratitude to my lab mates, from whom I have imbibed invaluable lessons during my graduate studies. Special mention goes to Mohamadali Morshedi, whose kind demeanor and interactions have been particularly enriching. My roommates and friends at Purdue have made being far from home more manageable. I extend recognition to Jenny Ricksy, Diana Knecht, Bonnie Sondgeroth, Cole Stonebraker, LeAnne Williams, and Katie Maish for their countless assistance throughout these five years.

Finally, I am deeply grateful for the constant and unconditional support of my parents, my sister, and my wife. Their numerous sacrifices have contributed immeasurably to my success, and I hold their roles in my journey in the highest esteem. My journey would not be the same without them.

TABLE OF CONTENTS

LIST OF TABLES 9
LIST OF FIGURES 10
ABSTRACT 12
1. INTRODUCTION 13
1.1 Research Background and Needs 15
1.2 Research Statement 18
1.3 Research Objectives 19
1.4 Research Scope 20
1.5 Research Data 21
1.6 Research Methodology 23
1.7 Expected Outcomes 25
1.8 Dissertation Organization 26
2. LITERATURE REVIEW 28
2.1 Historic Flood Losses 28
2.2 Flood Loss Estimation 31
2.3 Flood Risk Reduction 37
2.4 National Flood Insurance Program 43
2.5 Flood Risk Factors 56
2.6 Summary and Point of Departure 61
3. POST-DISASTER FEDERAL ASSISTANCE AND FLOOD INSURANCE DEMAND .. 64
3.1 Introduction 64
3.2 Research Background 66
3.3 Research Data 70
3.3.1 Treatment Variable 71
3.3.2 Outcome Variables 72
3.3.3 Confounding Variables 75
3.4 Research Methods 77
3.4.1 Part I: Propensity Score Matching for Binary Treatment 78
3.4.2 Part II: Generalized Propensity Score for Continuous Treatment 79

3.5 Results ... 81
3.5.1 Outcome for Binary Treatment ... 81
3.5.2 Outcome of Continuous Treatment ... 84
3.6 Discussion ... 89
3.7 Conclusion ... 91
4. FLOOD RISK FACTORS AND FLOOD INSURANCE PAYOUT ... 92
4.1 Introduction ... 92
4.2 Research Background ... 94
4.3 Flood Risk Factors ... 97
4.3.1 Flood Exposure ... 97
4.3.2 Infrastructure Vulnerability ... 98
4.3.3 Social Vulnerability ... 100
4.3.4 Community Resilience ... 101
4.3.5 Mobile Homes ... 103
4.4 Research Data and Methods ... 104
4.4.1 Response Variable ... 104
4.4.2 Control Variables ... 107
4.4.3 Research Methods ... 109
4.5 Results and Discussions ... 110
4.6 Conclusion ... 116
5. PREDICTING ANNUAL FLOOD INSURANCE PAYOUT ... 118
5.1 Introduction ... 118
5.2 Research Background ... 120
5.3 Research Data and Methods ... 123
5.3.1 Predictor Variables ... 124
5.3.2 Research Methods ... 131
Ordinary Least Square Regression (OLS) ... 132
Robust Regression (RR) ... 132
Generalized Linear Model (GLM) ... 134
5.4 Results ... 134
5.5 Conclusion ... 140

6. CONCLUSION 142
6.1 Summary and Contributions to the Body of Knowledge 142
6.2 Research Limitations 148
6.3 Recommendations for Future Research 150
REFERENCES 152
VITA 171

LIST OF TABLES

Table 1.1 Sources of the Collected Data 22
Table 2.1 Summary of NFIP Reinsurance Placements (Source: FEMA) 38
Table 2.2 Summary of NFIP CAT Bond Placements (Source: FEMA) 39
Table 2.3 List of Flood Risk Factor 58
Table 3.1 Previous Research on Charity Hazard in the Flood Insurance Market 68
Table 3.2 List of All Variables along with Descriptive Statistics ($t \in [2016, 2019]$) 73
Table 3.3 Standardized Mean Difference Before and After Matching 82
Table 3.4 Intergroup Comparison of Outcome Variables 83
Table 3.5 Regression Coefficients of OLS and WLS for Dose Response Function 87
Table 4.1 List of Variables with Descriptive Statistics 105
Table 4.2 Regression Coefficients for LMM 111
Table 5.1 List of Variables with Descriptive Statistics on Training Set 126
Table 5.2 Regression Coefficients for Initial Models 135
Table 5.3 Predictive Performance of Initial Models on the Training Set 136
Table 5.4 Predictive Performance of Final Models on Training Set 137
Table 5.5 Predictive Performance of Final Ensemble Model on Test Set 137
Table 5.6 State wise Comparison of Actual and Predicted Annual NFIP Payout in 2021 139

LIST OF FIGURES

Figure 1.1 Change in Number of Properties Located in 100-year Flood Zones 14

Figure 1.2 Average Flood Insurance Penetration Rate after Major Disasters 16

Figure 1.3 Methodological Framework 24

Figure 2.1 Comparison of Disaster Events by Type: 1980 – 1999 vs. 2000 – 2019 (Source: UNDRR 2020) 29

Figure 2.2 Historical Losses from Floods and Storms in the U.S. 30

Figure 2.3 NOAA Billion-dollar Events (Source: NOAA 2022) 31

Figure 2.4 Types of Housing Assistance and Other Needs Assistance (Source: Webster 2019) . 53

Figure 2.5. Conceptual Framework for Long-Term Sustainability of the NFIP 62

Figure 3.1 Spatial Distribution of the Counties Used in the Analysis 71

Figure 3.2 Correlation Matrix of the Covariates 77

Figure 3.3 Research Framework for Binary Treatment 79

Figure 3.4 Research Framework for Continuous Treatment 80

Figure 3.5 Comparison of Outcome Variables 83

Figure 3.6 Test for Balancing of Covariates for the IHP Count 86

Figure 3.7 Test for Balancing of Covariates for the IHP Amount 86

Figure 3.8 Dose Response Function for IHP Count 88

Figure 3.9 Dose Response Function for IHP Amount 89

Figure 3.10 Boxplots of Percentage Eligibility and Average IHP Amount 90

Figure 4.1 Flood Exposure of Counties (Log Transformed) 98

Figure 4.2 Per Capita Public Assistance Payout in Counties (Log Transformed) 100

Figure 4.3 Average Annual SVI of Counties 101

Figure 4.4 Community Resilience of Counties 102

Figure 4.5 Correlation Matrix 108

Figure 4.6 Residual Plots. (a) Histogram (b) Q-Q Plot 116

Figure 5.1 Spatial Distribution of Counties Analyzed in this research 123

Figure 5.2 Correlation Matrix 130

Figure 5.3 Research Framework 131

Figure 5.4 Scatter Plot of Actual vs Predicted Annual NFIP Payout on Test Set 138

ABSTRACT

The National Flood Insurance Program (NFIP), which is run by the U.S. Federal Emergency Management Agency (FEMA), is presently under huge debt to the U.S. treasury. The debt is primarily caused by low flood insurance take-up rate, low willingness to pay for flood insurance, and large payouts after major disasters. Addressing this insolvency problem requires the NFIP to understand (1) what drives the demand for flood insurance so that it can be increased, (2) how risk factors contribute towards large flood insurance payouts so that effective risk reduction policies can be planned, and (3) how to predict the future flood insurance payouts so that the NFIP can be financially prepared. This research has answered these three fundamental questions by developing empirical models based on historical data. To answer the first question, this research has developed a propensity score-based causal model that analyzed one of the key components that influences the demand for flood insurance – the availability of post-disaster government assistance. It was found that the availability of the federal payout in a county in a year increased the number of flood insurance policies by 5.2% and the total insured value of the policies by 4.6% in the following year. Next, this research has developed Mixed Effects Regression model that quantified the causal relationships between the annual flood insurance payout in a county and flood related risk factors such as flood exposure, infrastructure vulnerability, social vulnerability, community resilience, and the number of mobile homes in the county. Based on the derived causal estimates, it was predicted that climate change, which is expected to increase flood exposure in coastal counties, will increase the annual NFIP payout in New Orleans, Louisiana by $2.04 billion in the next 30 years. Lastly, to make the NFIP financially prepared for future payouts, this research has developed a predictive model that can predict the annual NFIP payout in a county with adequate predictive accuracy. The predictive model was used to predict the NFIP payout for 2021 and it was able to predict that with a 9.8% prediction error. The outcomes of this research create new knowledge to inform policy decisions and strategies aimed at fortifying the NFIP. This includes strategies such as flood protection infrastructure, tailored disaster assistance, and other interventions that can bolster flood insurance uptake while mitigating the risk of substantial payouts. Ultimately, this research contributes to sustaining the NFIP's ability to provide vital flood insurance coverage to millions of Americans.

1. INTRODUCTION

Natural hazards cause extensive damage to human life, infrastructure, property, economy, etc. Historical data from Munich Re shows that the frequency of natural hazards has increased steadily since 1980. Since then, disasters due to natural hazards have caused a cumulative loss of \$5.2 trillion globally (Munich Re 2020). Among the natural hazards, the losses due to floods are by far the highest on a global scale (Colgan et al. 2017, Dubbelboer et al. 2017, CRED-UNISDR 2015). The U.S. is no exception to that (Munich Re 2020). In the U.S., the cumulative loss from floods between 1988 and 2017 has been \$199 billion (Davenport et al. 2021).

Quinn et al. (2019) have analyzed 40 years of historical flood data and have found that there is a 1% chance of the losses from fluvial (river) floods exceeding \$78 billion in any given year in the U.S. Additionally, there is a 0.1% chance of the cost exceeding \$136 billion. Armal et al. (2020) have stated that the direct flood losses in the U.S. have risen from \$4 billion annually in 1980 to \$17 billion annually between 2010 and 2018. Jevrejeva et al. (2018) have forecasted that the global flood loss can exceed an additional \$1.4 trillion annually if the rise of global temperature is not maintained at 1.5 °C and reaches 2 °C. This can potentially cause an increase of the global sea level by an additional 11 cm. Due to global warming the precipitation extremes have changed across many regions in the U.S. Davenport et al. (2021) has utilized historical flood damage data and found that between 1988 and 2017, the cumulative impact of the precipitation change has been \$73 billion. Wing et al. (2022) estimated a \$32.1 billion annual loss from floods in the U.S. based on 2020's climate scenarios. They also found that the flood losses are borne disproportionately by the poorer communities.

In 2020, the 1st Street Foundation published a report that estimated the number of properties that are located in 100-year flood zones (1st Street Foundation 2020). Their estimation of the number of properties situated in 100-year flood zone was around 1.7 times compared to the U.S. Federal Emergency Management Agency (FEMA) 100-year flood zone designation. They reported that approximately 40% of the homeowners who live in the 100-year flood zone are currently uninformed of or miscalculating the flood risk they face because they are not classified as being within the 100-year flood zone by FEMA. Figure 1.1. shows the expected changes in the number of properties located in the 100-year flood zone in 20 major U.S. cities with the highest flood risk based on the model developed by 1st Street Foundation (2020). It can be noticed that in all twenty cities the number of properties within a 100-year flood zone will increase by 2050, due to climate change.

Figure 1.1 Change in Number of Properties Located in 100-year Flood Zones

To mitigate the risk of natural hazards, countries use different risk finance strategies. They can be ex-ante, ex-post, or the combination of the two. Ex-ante strategies include risk transfer instruments like insurance, reinsurance, catastrophe bonds, and ex-ante budget allocations whereas ex-post strategies consist of contingent credit, emergency ex-post budget allocation, and ex-post direct credits like borrowings and post-disaster debts (World Bank Group 2016). Though the sustainability of ex-post strategies has been questioned (Lester and Gurenko 2004), the World Bank recommends a layered combination of different strategies to be the most efficient (Gurenko and Mahul 2003, Ghesquiere and Mahul 2010, GFDRR and World Bank Group 2014). Although ex-ante risk transfer instruments such as insurance are available in many countries, there is a vast gap between total economic and insured losses. Data from Munich Re show that more than 70% of the estimated $5.2 trillion global loss from floods since 1980 were uninsured (Munich Re 2020). Governments are often referred to as “the insurer of last resort” since they are assumed to accomplish the requirements of filling up this gap in finances after catastrophes. At micro level, countries and governments use risk sharing instruments like flood insurance where businesses and homeowners insure themselves from possible flood losses by purchasing flood insurance from government or private organizations. Different countries have different policies regarding flood insurance for their citizens. Although purchasing flood insurance are compulsory in some countries (Romania, Poland, Iceland, etc.), it is voluntary in most of the world (Atreya et al. 2015).

1.1 Research Background and Needs

The National Flood Insurance Program (NFIP), which is run by the Federal Emergency Management Agency (FEMA), started in 1968 under the National Flood Insurance Act. The reluctance of private insurers to provide flood insurance created the need for the NFIP (Kousky et al. 2020). Purchasing flood insurance was voluntary till 1973. After that, buying flood insurance was mandated for properties with mortgage from a federally regulated or backed lender that are located in a NFIP participating community and within 100-year flood zone by the Flood Disaster Protection Act of 1973. While the Federal government offers flood insurance to the households, it has been observed in the aftermath of the flood related disasters that most of the sufferers are uninsured or underinsured (Kousky 2011). A congressional research report published in 2019, Horn (2019), showed the flood insurance take up rate for some of the recent flood events, which is shown in Figure 1.2. It recorded the average NFIP take up rate in the Special Flood Hazard Area (SFHA), i.e., 100-year flood zones for multiple flood events such as the South Carolina Flood in 2015 (30%), Louisiana Flood in 2016 (31%), Hurricane Harvey in Texas (21%), Hurricane Irma in Florida (31%). It can be noticed that the average take-up rate across the counties, which cover more area than 100-year flood zones, is even lower than that of SFHA. Munich Re reported in 2020 that there were 14.6 million properties in the U.S. that were at substantial flood risk, i.e., located in 100-year flood zone. However, historical records show that in 2020, there were approximately 4 million active NFIP policies in the U.S., which also demonstrates the low take-up rate of flood insurance in the U.S. Moreover, it has also been found that NFIP policies are often short lived, i.e., they do not get renewed (Michel-Kerjan et al. 2012).

Due to the low flood insurance take-up rate, the U.S. federal government, as the insurer of last resort, compensates the disaster survivors who are underinsured and/or uninsured through FEMA managed Individual Assistance (IA) program. The Individuals and Households Program (IHP) within IA is the primary way FEMA supports disaster survivors (Webster 2019). IHP provides direct financial aid to qualified individuals and households who are underinsured or uninsured and have serious needs of support as a result of a presidentially declared emergency or a major disaster. To be eligible to receive IHP assistance, an applicant must furnish that (1) the damage is uninsured, (2) he or she is a citizen (or qualified alien), and (3) the property is the primary residence (Kousky and Shabman 2012).

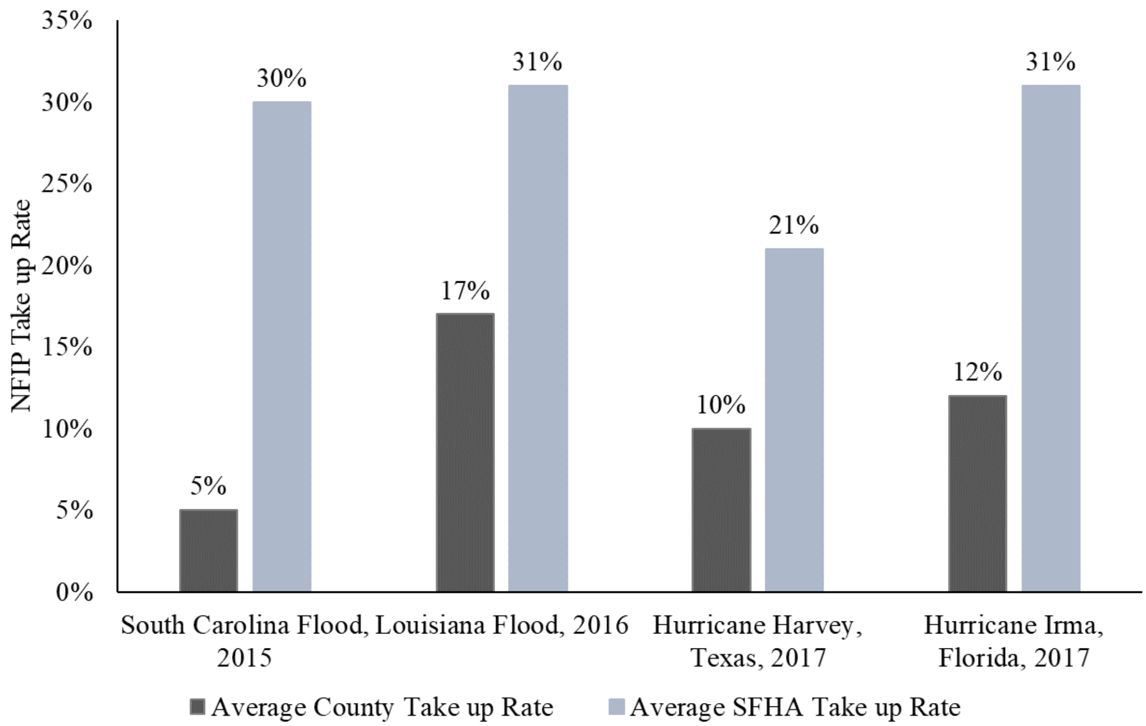


Figure 1.2 Average Flood Insurance Penetration Rate after Major Disasters

The demand for flood insurance is influenced by several factors such as the premium level, household income, damage from recent flood events, population, post-disaster federal support, etc. (Browne and Hoyt 2000, Kousky et al. 2018, Landry et al. 2021). There is a general consensus about how these factors influence the demand for flood insurance except for post-disaster federal support. It has been found that the availability of post-disaster government support often crowds out the demand for flood insurance, an event popularly known as Charity Hazard (Browne and Hoyt 2000, Raschky and Weck-Hannemann 2007). In the U.S., the federal regulations require that the IHP recipients maintain flood insurance for future assistance. Therefore, IHP should not crowd out the demand for flood insurance. Despite that, researchers have found conflicting evidence on the existence of charity hazard in the U.S. flood insurance market.

Understanding how the post-disaster federal assistance influences the flood insurance take-up rate in the U.S. is essential as it can significantly influence the cost of flood events to the federal government. If the IHP assistance crowds out flood insurance, then in future uninsured flood losses will greatly increase as more households will rely on the federal government instead of purchasing flood insurance. On the other hand, this can also increase the risk of large payouts to the NFIP because only the properties with high flood exposure might be left in the pool of NFIP due to presence of asymmetrically used information between insurer and insured. Bradt et al. (2021) have provided with strong evidence supporting the existence of asymmetrically used information in the NFIP. Furthermore, due to the increasing frequency of natural hazards, this domino effect of charity hazard could lead to further insolvency issues for the NFIP. Thus, this scenario increases the challenge of running both NFIP and IHP programs. Hence, it is essential to understand whether the Charity Hazard exists in the U.S. flood insurance market or not.

Additionally, NFIP had a cumulative debt of $20.5 billion to the U.S. Treasury as of 2020 after the federal government forgave $16 billion debt in 2016 (Grigg 2019, Horn 2020). The debt could be partly attributed to the low take-up rate. Moreover, it was projected that the annual deficit of collected premium and the expected payout would remain $1.4 billion in future (CBO 2017). This problem cannot be solved simply by raising the cost of insurance premiums. Previous researchers have found that the price elasticity of the demand for flood insurance is inelastic, which means that the demand for flood insurance is relatively insensitive to the price (Browne and Hoyt 2000, Landry and Jahan-Parvar 2011). However, with a higher cost of flood insurance premium, the demand for flood insurance might reduce further and NFIP might end up with the adverse selection problem, where only the households with high flood exposure purchase the flood insurance. The adverse selection problem, which arises due to the presence of asymmetrically used information between insurer and insured, is expected to increase the likelihood of future payouts. On the other hand, NFIP as a government sponsored program is the insurer of last resort even for the households that are deemed uninsurable by private flood insurers (FEMA 2015, Horn and Webel 2021). This restricts NFIP's ability to cherry pick households to reduce the likelihood of future payouts. Under this circumstances, it is essential that NFIP plans for flood risk reduction.

To summarize, NFIP is faced with two monumental challenges from climate change induced increased frequency and severity of natural hazards. First, the low flood insurance take-up rate due to which, the uninsured losses from future floods with potentially increased severity is expected to increase. Second, the increased likelihood of more frequent large payouts caused by potentially increased frequency and severity of future floods that might engender insolvency of NFIP. The rising cost of running the NFIP and the potential of increased uninsured and/or underinsured flood losses presents an overwhelming fiscal challenge to the U.S. federal government, already burdened with a debt over $32 trillion in 2023 (U.S. Treasury 2023). This research has tried to address that challenge by answering three fundamental questions (1) how the flood insurance take-up rate can be improved (2) how future payouts can be predicted, and (3) how future payouts can be reduced through flood risk reduction.

1.2 Research Statement

This research is conducted to assist the U.S. federal government in planning strategies and/or policies to improve the flood insurance penetration and keep the program solvent despite the increased intensity of natural hazards in future. As explained earlier, the solution cannot be simply achieved by increasing the flood insurance premium as it might increase the likelihood of large payouts due to adverse selection problems. On the other hand, decreasing the insurance premium to increase the flood insurance demand without any additional measure will reduce the revenue generated due to the existing inelastic relationship between price and demand for flood insurance (Browne and Hoyt 2000, Landry and Jahan-Parvar 2011). Reduced revenue while increasing intensity of natural hazards might not keep the program solvent in future. Under this scenario, it is essential to plan for flood risk reduction to keep the NFIP solvent in the long term so that its existence is not questioned and more importantly it can continue to fulfill its aim to mitigate the effect of flooding by providing affordable flood insurance. Additionally, through flood risk reduction if the government can reduce the likelihood of large payouts, it can also reduce the cost of flood insurance premium for the policy holders, which might increase the flood insurance take-up rate. Based on the research question and the proposed solution, the thesis statement is formed as below.

<u>To mitigate flood risk through flood insurance, it is essential to increase the flood insurance take-up rate and simultaneously decrease the likelihood of large payouts. This can only be achieved</u>

through flood risk reduction. Effective flood risk reduction policies can be more efficiently planned if the decision makers have the knowledge of the contributing factors of flood insurance demand and payouts based on data driven robust models.

1.3 Research Objectives

As explained in the previous sections, this research aims to answer three fundamental questions regarding the NFIP (1) how the flood insurance take-up rate can be improved (2) how future payouts can be predicted, and (3) how future payouts can be reduced through flood risk reduction. These three research questions explain the major research objectives. They are as follows.

- Objective 1 – Understand the factors that impact the demand for flood insurance in the U.S., particularly the availability of post-disaster federal assistance in terms of IHP payouts. It is important to understand how the availability of the IHP assistance influences the demand for flood insurance because it can greatly influence the cost of future floods to the U.S. federal government as explained previously. The outcomes of objective 1 can help in designing tailored disaster assistance policies to increase the demand for flood insurance in the U.S.
- Objective 2 – Quantify the causal relationship between different flood related risk factors such as flood exposure, infrastructure vulnerability, social vulnerability, etc., and flood insurance payout in the U.S. The causal relationships derived in this research will be helpful in planning for effective flood risk reduction measures that can keep future flood insurance payouts under control. Additionally, it will also help in estimating the benefits from different flood risk reduction initiatives such as improving infrastructure resilience, property buyout, etc. The outcomes of objective 2 can help in designing risk reduction strategies and/or policies that can reduce future payouts after large disasters.
- Objective 3 – Develop a model that can predict the annual NFIP payout with adequate predictive accuracy. The proposed model can be used to estimate the expected NFIP payout in different future climate scenarios, which can help in testing different counterfactuals. Moreover, NFIP payouts reflect the insured flood losses. If insured flood losses can be predicted, it can be useful to predict the uninsured flood losses as well, which can further be used to estimate the expected post-disaster IHP assistance. The outcome of objective 3 can ultimately help the NFIP to be better prepared for future payouts.

Once these objectives are achieved, this research will be able to answer the identified research questions and subsequently provide pathways to improve the flood insurance penetration and reduce the likelihood of large future payouts to keep NFIP solvent in the long term.

1.4 Research Scope

The successful completion of the research by fulfilling the research objectives requires defining specific research scope. First, the models will be developed at a macro level. It has been explained before that the research aims to solve the insolvency problem to ensure long-term sustainability of the NFIP. Therefore, the problem has been investigated from the perspective of the primary insurer, i.e., NFIP, which is also the same entity as the government. Hence, macro level analyses were deemed more appropriate for this research.

In terms of spatial scope, this research will concentrate on the 50 U.S. states and the District of Columbia. The analysis will not consider the U.S. territories such as Puerto Rico, Guam, U.S. Virgin Islands, Northern Mariana Islands, American Samoa, etc., as some of the data were not available for these territories. Moreover, these territories are fundamentally different from the U.S. states in the sense that not all U.S. laws are applicable to the territories (Webber 2017). Thus, they are not included in the research.

The research questions require collecting historical data. Therefore, recent historical data between 2016 and 2021 will be collected and used in this research. The models will be developed at the county level (i.e., the spatial unit is county) using annual data (i.e., the temporal unit is year). So, for the Objective 1, the factors that influence the annual demand for flood insurance in county will be analyzed. For Objective 2, causal relationships between flood risk factors and annual flood insurance payout in a county will be quantified. Lastly, for Objective 3, the proposed model will be able to predict the expected annual flood insurance payout in a county.

For Objective 1, this research will only analyze the influence of post-disaster federal assistance on the annual flood insurance enrollment in a county. The other factors that influence the demand for flood insurance such as premium price, household income, previous experience with floods, etc., will not be analyzed as there is a general consensus among the researchers on how these factors influence the demand for flood insurance (Browne and Hoyt 2000, Kousky et al. 2018, Landry et al. 2021). Thus, those factors have not been included in the scope of this research. The post-disaster federal assistance will be quantified in terms of the annual IHP payout in a county.

It has been explained before that IHP assistance is the primary way FEMA supports disaster survivors who are underinsured or uninsured (Webster 2019). Hence, focus will be limited to the IHP assistance only.

For Objective 2, the research plans to quantify the causal relationships between flood related risk factors such as flood exposure, infrastructure vulnerability, social vulnerability, etc., and the annual flood insurance payout in a county. It is important to note that there is no exhaustive list of flood risk factors. As a result, this research will only focus on the risk factors that have appeared the most in previous literature and can be controlled through human interventions. This has led to five controllable factors that influence the extent of flood losses in a region. They are (1) flood exposure, (2) infrastructure vulnerability, (3) social vulnerability, (4) community resilience, and (5) the number of mobile homes in the county.

The third and final objective of this research is to develop a model that can predict the annual flood insurance payout with adequate accuracy. The model will be developed using five years of data between 2016 and 2020 and tested on 2021 data. While there is no agreement in the existing literature on acceptable predictive accuracy, models will be considered adequate if the percentage error is less than ten percent. To achieve that, different regression techniques will be adopted, and their results will be compared for identifying the best model that can fulfill the ten percent requirement.

1.5 Research Data

To answer the identified three research questions, this research collected historical data between 2016 and 2021 from several publicly available data sources such as Federal Emergency Management Agency (FEMA), National Oceanic and Atmospheric Administration (NOAA), U.S. Census Bureau, Centers for Disease Control and Prevention (CDC), etc., as shown in Table 1.1. All the collected data were aggregated at the county level to perform the analysis. It is important to note that some of the datasets shown in Table 1.1 were not available for six years between 2016 and 2021. In such circumstances, different assumptions were made, which have been discussed in detail in later chapters. Moreover, the data cleaning and preprocessing steps corresponding to each research question have also been explained in detail in chapter three, four, and five of this dissertation. This section only lists the sources of the collected data.

Table 1.1 Sources of the Collected Data

No.	Data	Data Source
1	NFIP Payout	FEMA's FIMA NFIP Redacted Claims dataset
2	IHP Payout	FEMA's Registration Intake and Individuals Household Program dataset
3	Public Assistance (PA) Payout	FEMA's PA Funded Project Details dataset
4	Precipitation	NOAA's National Centers for Environmental information
5	Flood Damage	NOAA's Storm Event Database
6	County Area	U.S. Census Bureau's American Community Survey
7	Population	U.S. Census Bureau's American Community Survey
8	Median Building Age	U.S. Census Bureau's American Community Survey
9	Median Building Value	U.S. Census Bureau's American Community Survey
10	Percentage Occupancy	U.S. Census Bureau's American Community Survey
11	Education Level	U.S. Census Bureau's American Community Survey
12	Labor Force Participation	U.S. Census Bureau's American Community Survey
13	Median Household Income	U.S. Census Bureau's American Community Survey
14	Federal Mortgage	U.S. Federal Housing Finance Agency
15	No. of Mobile Homes	U.S. Census Bureau's American Community Survey
16	Flood Exposure	FEMA's National Risk Index for Natural Hazards
17	Social Vulnerability	U.S. Centers for Disease Control and Prevention
18	Community Resilience	University of South Carolina's Baseline Resilience Indicators for Communities (BRIC) dataset (Cutter et al. 2014)
19	Flood Insurance Policies and Total Insured Value	FEMA's National Flood Insurance Program (NFIP) Reinsurance Placement Information

1.6 Research Methodology

Figure 1.3. shows the methodological framework that has been adopted in this research. It should be noted that each research question and objective have a specific research methodology that has been discussed in detail in the following chapters. This section provides a summary of the methodological framework. It starts with the review of the existing literature on the research topic. The reviewed literature covers multiple domains of research such as historic flood losses, flood loss estimation, data-driven flood risk quantification, flood insurance, federal assistance program, charity hazard, flood risk reduction, etc. The reviewed literature facilitated formulating the research questions and the scope. Next, relevant data was collected. As explained previously, data was collected for six years between 2016 and 2021 from different publicly available data sources like U.S. Federal Emergency Management Agency's open data portal, the National Oceanic and Atmospheric Administration's database, U.S. Census Bureau, U.S. Centers for Disease Control and Protection, etc. All the collected data was cleaned and aggregated at the county level for the analysis.

The first research objective tests the existence of charity hazard in the U.S. flood insurance market. To be more specific, the first objective answers two questions (1) how the NFIP enrollment differs between the counties that received IHP payout and the counties that did not receive the IHP payout despite the declaration of major flood related disaster thus making them eligible to receive IHP assistance and (2) how different levels of IHP payout influences the NFIP enrollment in a county. To answer these questions, this research has used the collected data in a Propensity Score Matching (PSM) method to answer the first research question and Generalized Propensity Score (GPS) method to answer the second research question. The outcome of objective 1 infers whether the IHP assistance crowds out the demand for NFIP or not.

In the second objective, this research has developed an empirical model that quantifies the causal relationship between annual flood insurance payout and different flood related risk factors such as flood exposure, infrastructure vulnerability, social vulnerability, community resilience, number of mobile homes, etc., by using historical data for six years between 2016 and 2021 in a Linear Mixed Effects Regression model. Although it is known that the identified flood risk factors influence the flood insurance payout, the proposed model quantifies that causal relationship while considering necessary control variables.

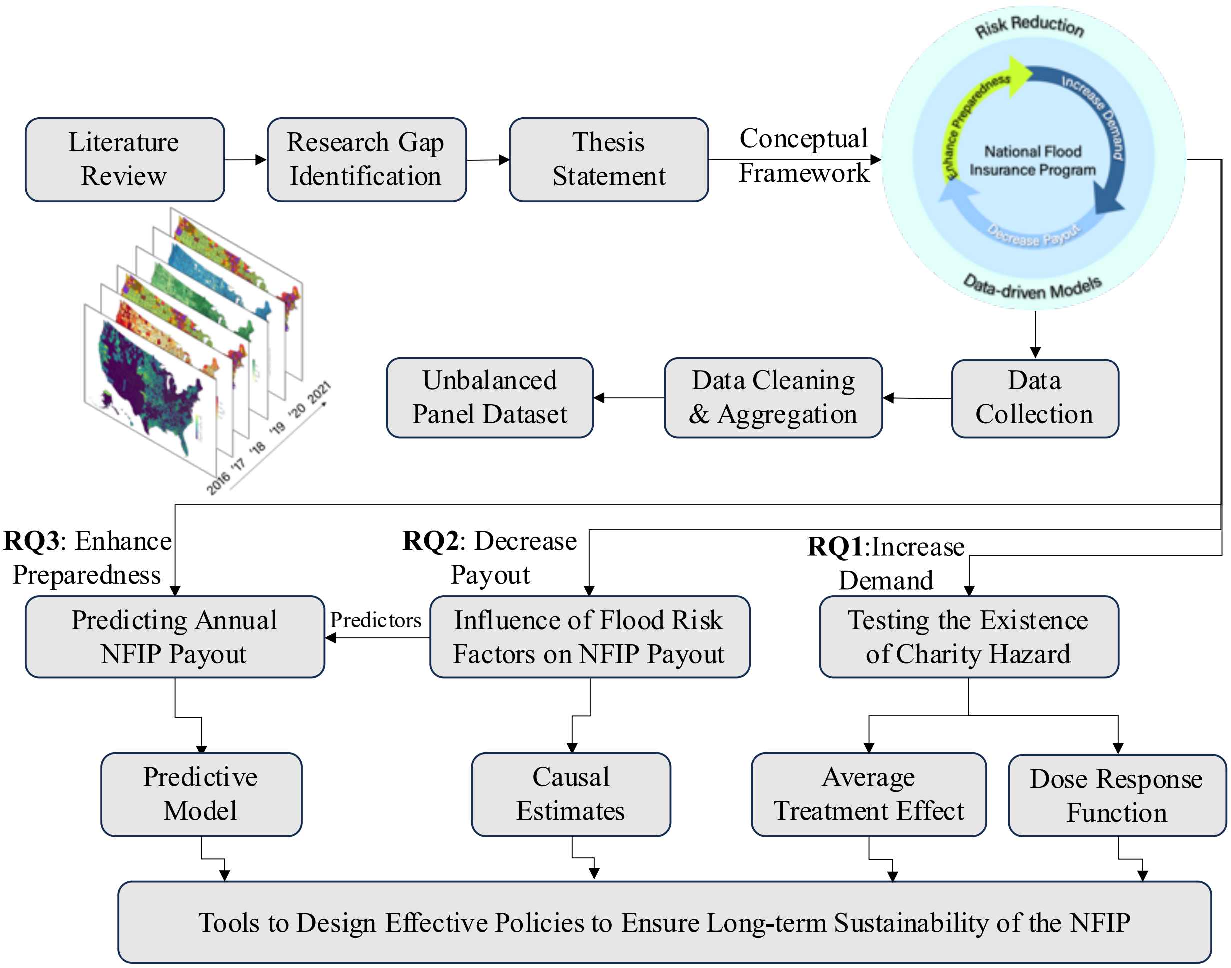


Figure 1.3 Methodological Framework

In Objective 3, a prediction model that can predict the annual NFIP payout in a county based on different factors such as flood damage, flood exposure, infrastructure vulnerability, social vulnerability, community resilience, number of NFIP policies, number of NFIP claims, total insured value, etc., has been developed using five years of historical NFIP payout data between 2016 and 2020 by adopting three regression techniques (1) Ordinary Least Square Regression, (2) Robust Regression, and (3) Generalized Linear Model. The fourth and the final ensemble model uses the outcomes of these three regression models to generate its own predictions. The ensemble model was used to predict the annual flood insurance payout of the flood affected counties for the year 2021.

1.7 Expected Outcomes

The expected outcomes of the research are listed below.

- First, the research will analyze the influence of IHP assistance on flood insurance enrollment. The outcome will reflect the difference between flood insurance enrollment in a county that received IHP assistance and a county that did not receive the IHP assistance. This difference will quantify the Average Treatment Effect (ATE) of the IHP assistance on flood insurance enrollment. ATE reflects the average difference between the treated and the non-treated groups. Next, the research will develop a dose response function that demonstrates the expected changes in flood insurance enrollment for different levels of IHP assistance. The dose response function will estimate the Average Treatment Effect on the Treated (ATT).
- Second, this research will quantify the causal relationship between flood risk factors and the annual flood insurance payout in a county. Although it is known that the identified flood risk factors affect the extent of flood damage and subsequently the extent of flood insurance claims, this research will quantify that influence. The outcome will be the coefficients that will explain the expected changes in the annual NFIP payout for a unit change in the flood risk factors.
- The derived regression equation that explains the causal relationships between the flood risk factors and the annual flood insurance payout while considering the required control variables can be used as an objective function for optimizing the annual flood insurance payout through the implementation of different flood risk reduction strategies and/or policies.
- Lastly, the proposed prediction model will be able to predict the annual flood insurance payout with adequate accuracy so that it can be used by the disaster management agencies in estimating the extent of flood insurance claims in a future year based on different climate scenarios. The proposed prediction model is expected to facilitate a stakeholder centric flood loss estimation. Like other natural hazards, flood losses are also shared among the stakeholders. Peng et al. (2014) and Wang et al. (2020) have listed four classes of stakeholders who are associated with losses from natural hazards. They are households, primary insurers, reinsurers, and governments. However, the flood loss is not necessarily shared equally among the stakeholders. There are several factors that influence this cost sharing such as government's policy regarding flood insurance, insurance penetration rate, risk transfer to reinsurance, etc. Therefore, a stakeholder-centric flood loss and flood risk assessment is more insightful than a generic one as it reflects the true cost of floods to each class of stakeholder. Although there is

plethora of flood loss and flood risk models, analysis on flood loss and flood risk from the perspective of a stakeholder is relatively underexplored. The proposed prediction is expected to be utilized to estimate flood loss from an insurer's perspective.

1.8 Dissertation Organization

This dissertation is organized in six chapters. The first chapter introduces the research background, needs, and statement. It also explains the research data, scope, methodology, and expected outcomes.

The second chapter explains the state of the art on this research topic and the point of departure from the existing body of knowledge. As explained previously, the scope of the research required review of existing literature on different subjects such as flood risk, flood insurnace, data-driven models, flood risk reduction, causal models, etc. They have been explained in detail in chapter 2.

The third chapter focuses on understanding the effect of one of the key factors that determine the demand for flood insurance in the U.S. The chapter tested the hypothesis that post-disaster IHP assistance affects the demand for flood insurance in the U.S. To do that, this research has conducted analysis to understand the causality between post-disaster federal assistance and flood insurance enrollment in the flood affected counties in the U.S. In the first part, the treatment variable has been considered as binary to compare the effect of the availability of federal assistance to that of non-availability of federal assistance in a flood affected county by using propensity score matching method. Next, the treatment variable was considered continuous and was used in a generalized propensity score method to develop a dose response function, i.e., a function that depicts the changes in the flood insurance enrollment based on different levels of IHP assistance.

The fourth chapter adresses the question on how to reduce the flood insurnace payout from future disasters. The thesis statement claims that flood risk reduction can reduce flood insurance payouts. To prove that, it is essential to establish the causal link between flood risk factors and flood insruance payout. Therefore, the fourth chapter presents a causal model that quantifies the causal relationship between flood risk factors and the flood insurance payout in the U.S. The flood risk factors that have been considered in this research are flood exposure, infrastructure vulnerability, social vulnerability, community resilience, and the number of mobile homes. Historical data for the annual flood insurance payout, flood risk factors, and other control variables were used in a Mixed Effects Regression model to derive the empirical relationships. The regression model expressed the natural logarithm of the annual flood insurance payout in a county based on the flood risk factors and control variables.

The fifth chapter adresses the final piece of the puzzle, i.e., how to enhance the financial preparedness of the NFIP for future payout. To do that, it is essential to have the ability to predict future payouts with a reasonable accuracy. Therefore, the fifth chapter presents a prediction model that can predict county level insured flood loss to households, measured in terms of annual flood insurance payout, in the U.S. based on different factors such as rainfall anomaly, flood damage, flood exposure, infrastructure vulnerability, social vulnerability, number of flood insurance policies, total insured value, etc. The prediction model has been developed using five years of historical flood insurance claims data between 2016 and 2020. For developing the model, three regression techniques were adopted (1) Ordinary Least Square Regression, (2) Robust Regression, and (3) Generalized Linear Model. The final ensemble model uses the outcomes of these three regression models to generate its own predictions. The ensemble model was used to predict the annual flood insurance payout of the flood affected counties for the year 2021.

Overall, this dissertation presents a comprehensive analysis of the Flood Insurance Program in the U.S., addressing critical challenges including low uptake rates and significant debt. Through causal modeling, risk factor assessment, and predictive modeling, the research provides actionable insights to strengthen the NFIP. The findings underscore the importance of tailored policies and interventions to increase flood insurance participation, reduce the likelihood of large payouts, and ensure the program's long-term financial viability for safeguarding millions of Americans. The final chapter of this dissertation concludes the research by summarizing the works presented in chapters three, four, and five, its contributions to the body of knowledge, limitations, and recommendations for future research. It is important to note that the second, third, and fourth chapters of this dissertation had been written in a journal article format. The articles were under review in different journals when this dissertation was written.

2. LITERATURE REVIEW

This research aims to develop models that can assist policy makers to increase the flood insurance take-up rate and decrease the likelihood of large payouts in future. Before the objectives can be achieved, it is essential to review the existing knowledge on the topic, which has been accomplished through the review of existing literature. The vast scope of the research required reviewing literature from multiple domains, which are: (1) Historic Flood Losses, (2) Flood Loss Estimation, (3) Flood Risk Reduction, (4) National Flood Insurance Program, and (5) Flood Risk Factors. This chapter explains the existing knowledge in these domains.

2.1 Historic Flood Losses

Natural hazards cause extensive damage to human life, infrastructure, property, economy, etc. Historic data from Munich Re show that the frequency of natural hazards has increased steadily since 1980. Since then, these natural hazards have caused a cumulative loss of $5.2 trillion globally (Munich Re 2020). The United Nations Office for Disaster Risk Reduction (UNDRR) published a report on October 13, 2020, the International Day for Disaster Risk Reduction, where it has displayed how extreme weather events like floods, severe storms, etc., have dominated the disaster landscape in the 21st century (UNDRR 2020). The data presented in that report is sourced from the Emergency Events Database (EM-DAT), managed by the Centre for Research on the Epidemiology of Disasters (CRED). EM-DAT records classified disasters as incidents where the death toll reaches ten or more, impacts a minimum of 100 individuals, leads to a declaration of state of emergency, or triggers a request for international aid. Figure 2.1 shows how the frequency of natural hazards have changed since the start of this century. Over the past two decades, there has been a significant increase in the occurrence of major floods across the globe, which has more than doubled from 1389 to 3254. Similarly, the frequency of storms has also risen from 1457 to 2034. Floods and storms have emerged as the most commonly observed events during this period as shown in Figure 2.1.

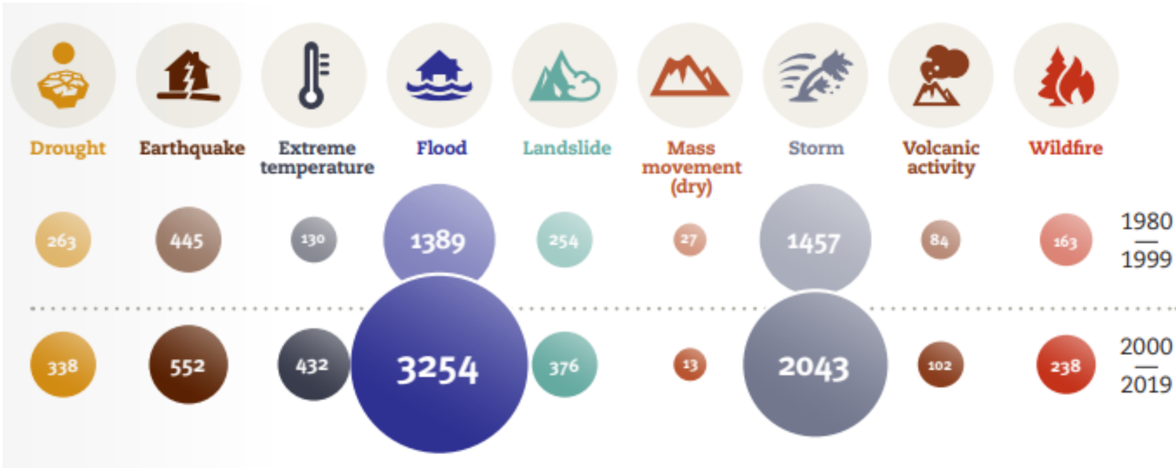


Figure 2.1 Comparison of Disaster Events by Type: 1980 – 1999 vs. 2000 – 2019 (Source: UNDRR 2020)

Among the natural hazards the losses due to floods are by far the highest on a global scale (Colgan et al. 2017, Dubbelboer et al. 2017, CRED-UNISDR 2015) and the U.S. is no exception to that (Munich Re 2020). In the U.S., the cumulative loss from floods between 1988 and 2017 has been \$199 billion (Davenport et al. 2021). Floods can be of different types. The National Oceanic and Atmospheric Administration's (NOAA) National Severe Storm Laboratory has categorized floods into five types. They are river flood, coastal flood, storm surge, inland flood, and flash flood. River floods take place when water levels exceed the riverbanks' capacity, often due to factors like prolonged rainfall, thunderstorms, snowmelt, and more. Coastal floods, on the other hand, result from exceptionally high tides exacerbated by heavy rainfall and winds blowing from the ocean toward land. Storm surges are characterized by an abnormal coastal water level rise, surpassing regular tidal patterns, propelled by the forces of a severe storm's wind, waves, and low atmospheric pressure. Inland floods transpire when moderate precipitation persists over days, there is an abrupt deluge in a short span, or a river breaches its boundaries due to obstructions like ice or debris, or due to dam or levee failures. Lastly, flash floods arise from intense rainfall in a brief timeframe, often less than six hours.

The EM-DAT database, which provides a comprehensive view of the historic losses resulting from different types of natural hazards, was created with the support from the World Health Organization (WHO) and Belgium Government. Figure 2.2 shows the historical trend in the losses from floods and storms since 1974 in the U.S. as recorded in the EM-DAT database.

Since the start of the century, the losses from the natural hazards such as floods and storms (which most often result in floods) have increased significantly.

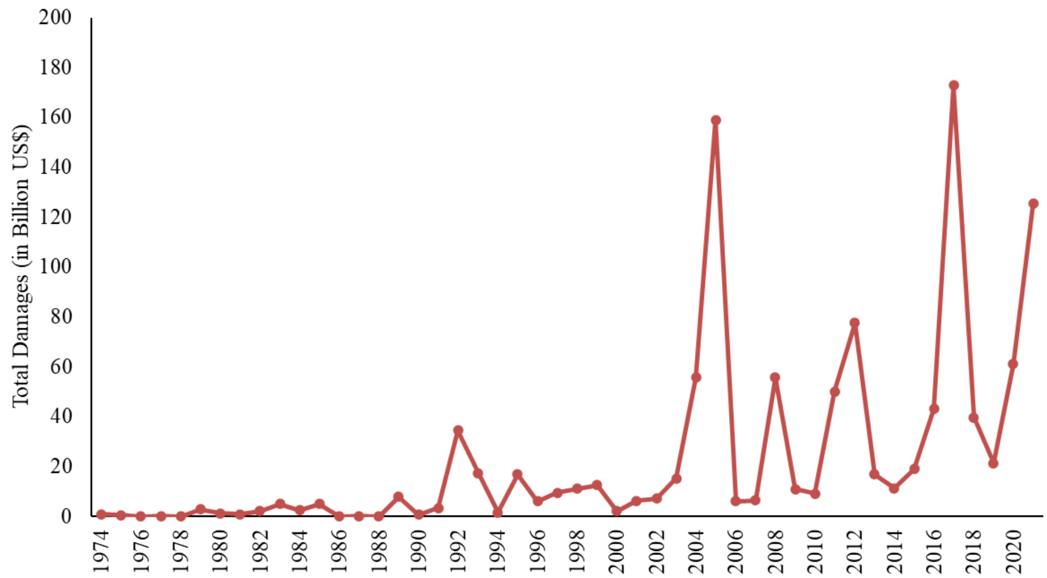


Figure 2.2 Historical Losses from Floods and Storms in the U.S.

NOAA maintains a record of disasters due to natural hazards in the U.S. where the cost exceeded a billion dollars and terms them as billion-dollar events. Figure 2.3 displays the cost and the frequency of these billion-dollar events. From the figure, it is apparent that the frequency of those events has also increased since the turn of the century. The cost of natural hazards has been increasing in the U.S. due to a combination of increase in exposure (more assets are exposed to hazards), vulnerability (susceptibility to failure to a natural hazard), and frequency due to climate change. Quinn et al. (2019) have analyzed 40 years of historical flood data and have found that there is a 1% chance of the losses from rive floods exceeding \$78 billion in any given year in the U.S. Also, there is a 0.1% chance of the cost exceeding \$136 billion. Armal et al. (2020) have stated that the direct flood losses in the U.S. have risen from \$4 billion annually in 1980 to \$17 billion annually between 2010 and 2018. Jevrejeva et al. (2018) have forecasted that the global flood loss can exceed an additional \$1.4 trillion annually if the rise of global temperature is not maintained at 1.5 °C and reaches 2 °C. This can potentially cause an increase of the global sea level by an additional 11 cm.

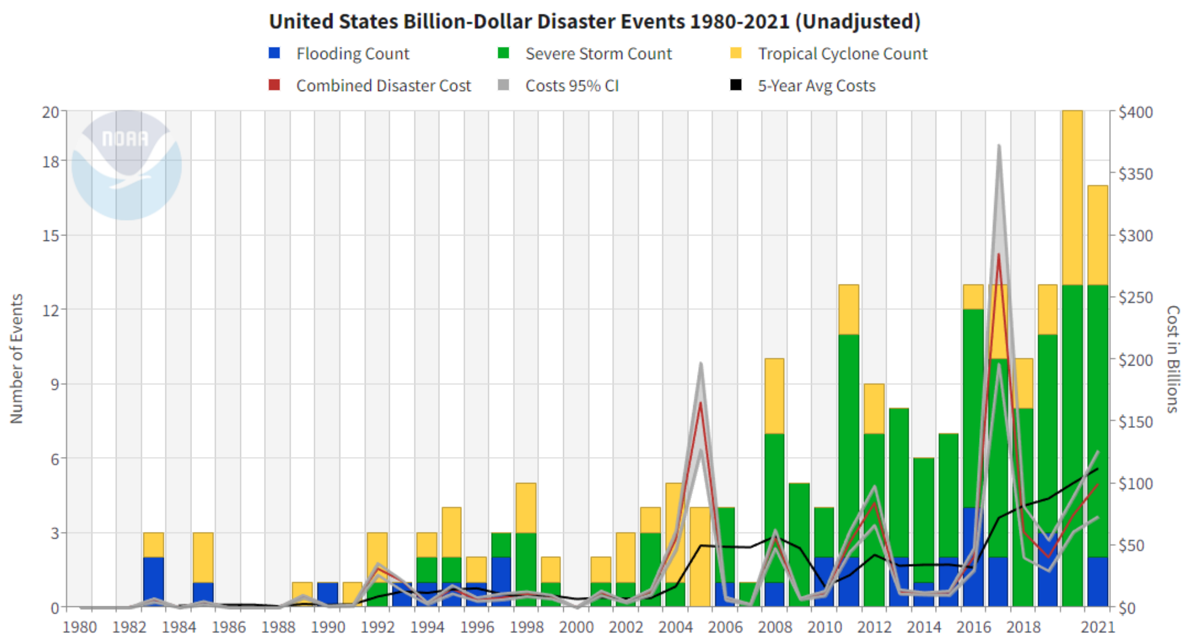


Figure 2.3 NOAA Billion-dollar Events (Source: NOAA 2022)

Due to global warming, the precipitation extremes have changed across many regions in the U.S. A recent study (Davenport et al. 2021) has utilized historical flood damage data and found that between 1988 and 2017, the cumulative impact of the precipitation change has been \$73 billion. Wing et al. (2022) have estimated a \$32.1 billion annual loss from floods based on 2020's climate scenarios.

2.2 Flood Loss Estimation

One popular approach for flood loss estimation is to utilize stage-damage function, which relates flood damage to various flood parameters (depth, duration, etc.) for different class of objects (Krzysztofowicz and Davis 1983, Smith 1994). Depth-damage functions are very popular in flood loss estimation research and have existed for more than two decades. Dutta et al. (2003) have them in estimating expected flood loss based on flood depth, velocity, and duration. The tangible damages were classified into three categories: urban (damage to buildings, structures, properties, etc.), rural (damage to agriculture products, farmhouses, etc.), and infrastructure (power, transportation, gas supply, etc.). Typically, the direct flood damage estimation to buildings comprises two interrelated steps (Pistrika and Jonkman, 2009). The initial step entails analyzing the structural damage resulting from flood impacts, which is influenced by flood actions and the building's resistance (Kelman and Spencer 2004). Subsequently, the economic estimation of physical damage is conducted. To convert the structural damage into monetary values, it is necessary to have knowledge of the pre-disaster market value and replacement cost of the property.

Arrighi et al. (2018) have measured flood risk at individual building level. They have used replacement and recovery costs for quantifying estimated damage to buildings. These costs varied for different buildings based on their usage. Rözer et al. (2019) have used historic flood loss data between 2005 and 2014 of five German cities to develop a prediction model to predict the degree of flood loss to the buildings based on water depth, duration, basement, contamination, and household size. Water depth and duration were found as the most important predictors. The FEMA has created a geographic information system (GIS) based application named HAZUS, that uses a depth-damage function to calculate the flood damage to different building types. The model can also estimate the expected tax losses from floods (Scawthorn et al. 2006).

Despite the popularity of depth-damage functions, there are certain limitations to them. These functions primarily consider water depth as the main element influencing direct damage However, additional factors such as flow velocity, duration of flooding, the presence of a flood warning system, and the efficiency of emergency response can also impact the level of flood damage to buildings (Pistrika and Jonkman 2009, Pistrika 2010, Tsakiris 2014). Unfortunately, most flood damage models do not incorporate all these factors. Second, they are site-specific. Therefore, the depth-damage function developed for one location may not be applicable for another location (Pistrika et al. 2014, Martínez-Gomariz et al. 2020).

On macro level, index-based flood vulnerability and resilience assessment has gained popularity in recent years. Ezell (2007) has defined vulnerability as the measure of proneness to threat scenarios like natural hazards, intentional attacks, etc. Balica et al. (2012) created a flood vulnerability index for cities based on three factors: exposure, susceptibility, and resilience (Balica et al. 2009). They have adopted a system-based approach in developing the index where they have identified three components that explain the flood vulnerability of a city: hydro-geological, socio-economic, and politico-administrative. For these three components they have identified 19 indicators: sea level rise, storm surge, number of cyclones, river discharge, foreshore slope, soil subsidence, coastline in km, cultural heritage (number of historical buildings in danger), population close to coastline, growing coastal population, shelters, percentage of disabled person, awareness and preparedness, recovery time, km of drainage, flood hazards map, institutional organizations, uncontrolled planning zones, and flood protection. However, they have not considered existing capacities that expedite the recovery process.

Karagiorgos et al. (2016) has defined flash flood vulnerability as a summation of physical and social vulnerability. For calculating physical vulnerability, an empirical estimate between the degree of loss and the intensity of the flash flood has been produced. The degree of loss is computed as the ratio of empirically collected loss and value of every property. The fitted distribution was used to quantify the physical vulnerability. However, their analysis has not considered the other types of losses like loss of business, cost of debris removal, etc. Yang et al. (2018) defined a flood vulnerability index built on exposure, sensitivity, adaptive capacity. Their list of indicators consists of flood velocity, water depth, flooded area, population sensitivity, economic sensitivity, agricultural sensitivity, early warning capabilities, and self-restoring capabilities. They have performed multiple flood vulnerability assessment by defining six flood vulnerability types and four levels of flooding.

Miguez and Veról (2017) have developed a flood risk index based on the flood characteristics (depth and duration) and effects (dwelling density, income per capita, and inadequate sanitation). The flood risk index is further used in developing the flood resilience index. The objective of the risk index is to compare different design options to mitigate urban floods. The resilience index basically estimates the efficiency of future projects in terms of the reduction of risk index. Chen and Leandro (2019) have proposed a time varying flood resilience index based on two stages of flooding: event phase and recovery phase. For the event phase, the indicator is a function of water depth, accumulated water depth, duration of flood, and rising rate of flood water. For the recovery phase the authors have included the social and economic factors like number of children, number of elderly people, and the income of the household along with the four factors of the event phase. Leandro et al. (2020) has used the flood resilience index proposed in Chen and Leandro (2019) to gauge the impacts of climate change adaptation measures in the households of Munich city. They have analyzed the impacts of flood proof gates and flood tanks in households.

More recently a mixed methodology of statistical methods, data analytics, and machine learning techniques have become more prominent in disaster risk reduction studies. Researchers have been using historical data to develop various types of empirical models to derive insights from those data. Luu et al. (2019) used a combination of Multiple Linear Regression (MLR) and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) to analyze Vietnam's regional flood risk based on their national disaster loss database (DANA). MLR was used to identify the weights for flood damage attributes. The TOPSIS method was then adopted the weights to rank the regions based on their flood risk. Pres (2009) also recommended the use of multivariate linear regression for enhanced quantification of weather risks for businesses. Varazanashvili et al. (2012) used multiple linear regression to identify the relationship between the financial losses, physical exposure, and six types of hazards to develop a national multi-risk map for the Republic of Georgia.

In addition to MLR, several machine learning techniques have been successfully used for vulnerability assessment and risk profiling of infrastructure systems. For instance, Leon and Atanasiu (2006) used k-nearest Neighbor graphs to develop a GIS-based clustering model for evaluating the seismic vulnerability of regions using a case study from Iasi, a large city in Romania. The provided clusters assist the decision makers to identify the buildings that belong to different risk or damage classes. Abdulla and Birgisson (2020) used different classification algorithms, e.g., k-nearest Neighbor (kNN), Random Forest, Logistic Regression, and Naive Bayes to classify the Cumulative Inundation (CI) of different nodes in the road network in Houston, Texas into three categories of not vulnerable, moderately, and highly vulnerable. They later translated the vulnerability of the road network resulted from removing a node due to fluvial flooding. Their results showed that kNN provided the highest prediction accuracy (Abdulla and Birgisson, 2020).

Mukherjee et al. (2018) used support vector machines and random forests algorithms to predict the state-level of power outages and categorize the risk factors using historical power outage data in the U.S. Although their framework can facilitate the investment and policy making at the federal level, the influence of the current state of the infrastructure system on its performance was not thoroughly investigated, primarily due to limited accessibility to detailed micro-level data on the power grid infrastructure highlighting the need for alternative methods.

Ma et al. (2021) have utilized historical individual assistance data in a four nested multi-level logistic regression model to find the likelihood of a household having homeowners' insurance based on individual income and community income inequalities. Cutter et al. (2014) have found that individual assistance is a significant predictor in predicting a community's post-disaster recovery. Wang and Sebastian (2021) have developed an empirical model to quantify flood vulnerability expressed as a percentage of property loss for a given water depth based on hazard distribution, property exposure, built environment conditions, socio-economic factors of a community using historical flood loss claim data. Helderop and Grubesic (2022) have used historical flood insurance payout data to derive different levels of hurricane vulnerability for in the state of Florida. Lim and Skidmore (2019) have used a Zero Inflated Negative Binomial (ZINB) regression to model the deaths from flood events in a U.S. county based on event specific, area specific, demographic, socioeconomic explanatory variables.

Wing et al. (2020) have used historical NFIP redacted claims data to derive several insights on flood depth-damage functions. They have found that the observed flood losses are a non-monotonic function of flood depth. They follow a beta function bimodal distributions for flood depth. Anbarasan et al. (2020) have used a convoluted deep neural network classification model to classify areas with chances and no chance of flood occurrence using variables such as water flow, water level, rain sensor, humidity etc. Their model found better prediction accuracy than other existing algorithms like artificial neural networks and deep neural networks.

Khosravi et al. (2018) have used decision tree algorithms to predict the flash flood susceptibility of Haraz watershed in Iran based on various predictors such as land use, ground slope, rainfall etc. In their analysis, alternating decision trees have emerged as the most suitable predictor. Costache (2019) has also used different decision tree algorithms to classify flash flood potential index in Romania. Although decision trees are primarily used for classification problems, they have also been used in regression problems in the past (Xu et al. 2005, Swetapadma and Yadav 2016, Pekel 2020, Rakhra et al. 2021, Zhang et al. 2021). Merz et al. (2013) have used decision tree regression model to predict the direct building damage from floods in Elbe and Danube catchments in Germany based on 28 predictors of five types: hydrologic, emergency measures, precaution and experience, building characteristics, and socio-economic status. Their analysis showed that tree-based models performed better than the traditional depth-damage models. Ragettli et al. (2017) have also used decision trees for detecting flash floods in ungauged mountain catchments in China. They have found that decision tree models outperformed the conventional models in detecting the probability of 10-year flood events by 20%.

Spekkers et al. (2014) have collected property insurance claim data between 1998 and 2011 in Netherlands and have used it in decision tree regression model to predict the average claim size, claim frequency after floods. They have found that frequency of insurance claim is influenced by rainfall intensity, real estate value, household income, etc. Darmawan et al. (2021) have also used decision tree regression to find the impact of total population on flood intensity. Their analysis found that decision tree model performed better than linear regression, polynomial regression, and ridge regression. Abedi et al. (2021) have implemented four different tree-based algorithms: classification and regression trees, random forest, boosted regression trees, and extreme gradient boosting. They have found that advanced tree-based algorithms performed better than the normal decision trees.

Lee et al. (2017) have used random forest and stochastic gradient boosting algorithm to develop flood susceptibility maps for Seoul metropolitan city in South Korea. In their analysis, random forest model performed better than the boosting algorithm in the validation set. Sadler et al. (2018) have used it for predicting the number of coastal flooding per storm event based on historical environmental data such as rainfall, tide, groundwater level, wind conditions, etc. In terms of prediction performance, the RF model has outperformed Poisson regression model. Liu et al. (2020) have utilized Whale optimization algorithm to optimize RF hyperparameters to develop a regression model that they have used to predict the flood resilience index of Jiansanjiang Administration of Heilongjiang Province of China. Schoppa et al. (2020) have used RF regression model for predicting the flood discharge of 95 study basins located in Canada and the U.S. They have found that the RF model's performance is competitive to the traditional hydrological models in predicting low and medium magnitude flood discharge. Desai and Ouarda (2021) have utilized several linear, nonlinear models in predicting the flood quantile of ungauged sites in Quebec Canada based on different predictors such as basin mean slope, annual mean total precipitation, annual mean degree days, etc., and found that Canonical Correlation Analysis based RF regression model performed better than linear regression and artificial neural networks models.

Ahmed and Lee (2021) have used Extreme Gradient Boosting (XGBoost) algorithm in predicting the flooding susceptibility of urban public transit systems of Toronto in Canada and have found that model can predict flooding susceptibility with more than 95% accuracy. Similar research was undertaken by Pham et al. (2021). They have utilized several boosting models such as adaptive boosting, boosted generalized linear model, XGBoost, deep boost to predict the flooding susceptibility in the Talar watershed, Mazandaran province, Iran. In their case the XGBoost algorithm produced a prediction accuracy of 87%. Chen et al. (2021) have also used XGBoost algorithm along with five other machine learning techniques in predicting the risk map of Pearl River Delta in Southern China based on twelve predictors that included rainfall, population density, soil type, road data, etc. In their case XGBoost algorithm produced a prediction accuracy of over 96%. Sanders et al. (2022) have developed a XGBoost based Flood Alert System (FAS) and found the algorithm's performance promising in predicting the rising limbs and timings of critical stages.

To summarize, there has been a shift in approach from traditional depth-damage models to multivariate prediction models in flood damage estimation in recent times. Wagenaar et al. (2017) have found that data-driven regression models reduced the mean absolute error of predicting the flood damages to residential buildings and contents by 20% when compared to traditional approaches like depth-damage function. Amadio et al. (2019) have also concluded that when extensive data are available to characterize flood events, multivariate models provide more reliable damage estimates than expert-based damage models for residential buildings. Wagenaar et al. (2018) claimed that multivariate prediction models with heterogenous data from multiple locations and flood events have high potential for developing improved flood damage estimates. Similar conclusion can be found in Kellermann et al. (2020), Schoppa et al. (2020).

2.3 Flood Risk Reduction

Flood risk reduction initiatives can be of mitigation type such as floodwalls/seawalls, floodgates, levees, evacuation routes, relocating people and property from flood exposed areas, elevated structures, property buyouts, etc. They can also be of adaptation type such as early warning systems, risk-based land use planning, nature-based solutions, social safety, and risk financing instruments (Jongman 2018). For the last two decades, the combination of adaptation and mitigation in flood risk planning has been widely discussed (Bizikova et al. 2007, Harvey et al. 2014, Locatelli et al. 2016, Grafakos et al. 2019). Hinkel et al. (2013) have concluded that adaptation and mitigation are complementary to each other. They have also stated mitigation measures can significantly reduce flood losses in less wealthy countries where annual flood losses are significant with respect to the gross domestic product.

Transferring the flood risk to private reinsurers or to capital market through reinsurance and/or catastrophe bonds can also be considered as risk reduction strategies. FEMA uses these two mechanisms to transfer some of the financial risks of the National Flood Insurance Program. Reinsurance plays a crucial role in the risk management strategies of insurance companies, acting as a form of insurance for insurers themselves. By paying premiums to reinsurers, insurance providers like the National Flood Insurance Program (NFIP) can mitigate their financial exposure and transfer the risk of large losses. This risk transfer mechanism provides a safety net, similar to the way flood insurance protects a home. Table 2.1 shows the summary of the reinsurance placement.

Table 2.1 Summary of NFIP Reinsurance Placements (Source: FEMA)

Year	Amount of Risk Transferred	Number of Reinsurers	Annual Premium Paid
2017	$1.042 billion	25	$150 million
2018	$1.46 billion	28	$235 million
2019	$1.32 billion	28	$186 million
2020	$1.33 billion	27	$205 million
2021	$1.153 billion	32	$195.8 million
2022	$1.064 billion	28	$171.9 million
2023	$.5025 billion	18	$90.2 million

Reinsurance is commonly utilized by private insurance companies worldwide as a risk management tool. Public entities also purchase reinsurance, with several U.S. states operating insurance providers that leverage reinsurance, including the Citizens Property Insurance Corporation of Florida, the California Earthquake Authority, and the Texas Windstorm Insurance Association. Different types of reinsurance exist, and the NFIP utilizes "property catastrophe per occurrence excess of loss" reinsurance. Under this arrangement, the reinsurer reimburses the insurer for losses that exceed a predetermined deductible. This form of reinsurance serves as a vital safeguard against losses stemming from natural disasters and other catastrophic events. To manage the NFIP flood risk, FEMA employs two reinsurance strategies: conventional reinsurance with one-year term and Insurance-Linked Securities (ILS) reinsurance involving three-year catastrophe bonds. The ILS reinsurance expires after three years unless a reinsurance claim depletes the coverage. FEMA started the traditional one-year reinsurance from January 2017 and as shown in Table 2.1, it paid $1.042 billion of the $9.03 billion of the annual NFIP claims in 2017 after hurricane Harvey. FEMA plans to further expand the NFIP Reinsurance Program and actively explores additional avenues to enhance protection against future flood losses.

A catastrophe bond (CAT) is a high-yield debt instrument, whose purpose is to raise money for firms in the insurance sector in the event of a natural hazard. A CAT bond permits the bond issuer to receive payments from the bond only if specific conditions, such as a natural hazard occurs, and the loss exceeds a predetermined value. In the event that a bond-protected occurrence triggers a payout to the insurance company, the responsibility of the issuer to pay interest and reimburse the principal is either postponed or entirely waived. FEMA started CAT bonds program in 2018. Through reinsurance and CAT bonds, the program plans to achieve cost reduction in risk transfer, gain access to greater market capacity, and enhance risk diversification through collaboration with various partners. Table 2.2 shows the summary of CAT bonds placed by NFIP.

Table 2.2 Summary of NFIP CAT Bond Placements (Source: FEMA)

Year	Duration (3 Years)	Premium Paid	Coverage Losses
2018	Aug. 1, 2018 - July 31, 2021	$62 million	3.5% of losses within $5 and $10 billion 13% of losses within $7.5 and $10 billion
2019	April 17, 2019 - April 16, 2022	$32 million	2.5% of losses within $6 and $8 billion 12.5% of losses within $8 and $10 billion
2020	Feb. 20, 2020 - Feb. 19 2023	$50.28 million	3.33% of losses within $6 and $9 billion 30% of losses within $9 and $10 billion
2021	Feb. 23, 2021 – Feb. 22, 2024	$79.44 million	12.5% of losses within $6 billion and $7 billion and 22.5% of losses within $7 billion and $9 billion.
2022	Feb. 23, 2022 - Feb. 22, 2025	$61.23 million	2.5% of losses within $6 billion and $7 billion 5% of losses within $7 billion and $9 billion and 32.5% of losses within $9 billion and $10 billion.
2023	Mar. 7, 2023 – Mar. 7, 2026	$50.37 million	5% of losses within $7 and $8 billion 11.25% of losses within $8 and $10 billion

The scope of the CAT bonds encompasses the 50 states, the District of Columbia, Puerto Rico, and the U.S. Virgin Islands. Since the NFIP has limited policies in other territories, flood claims in those areas can be adequately addressed without reliance on these placements. Hence, the geographic scope excludes American Samoa, Guam, and the Northern Mariana Islands. Furthermore, the capital market reinsurance placements specifically cover the risk associated with "named storms," which are storms or storm systems classified as tropical cyclones, tropical depressions, tropical storms, or hurricanes by the National Weather Service's National Hurricane

Center. Consequently, these CAT bonds apply to floods directly or indirectly caused by such named storms. It is important to note that major flood events that are not categorized as named storms would not trigger the three placements between 2021 to 2023 as shown in Table 2.2. However, analyzing catastrophe models and historical data reveals that the majority of the NFIP's risk concerning single flood events resulting in losses exceeding $6 billion is primarily associated with named storms. Lastly, each of the 2018, 2019, and 2020 CAT bond coverage terminated without a covered event occurring before the expiration date.

Although the reinsurance along with the CAT bonds provides significant coverage of possible losses, they might still not be sufficient. For instance, the Caribbean Catastrophe Risk Insurance Facility (CCRIF) which is a multi-country risk pool, provides coverage to 22 Caribbean and Central American countries (CCRIF 2021a). The CCRIF disbursed $12.8 million in 2019 after cyclone Dorian to the government of Bahamas and $30.6 million in 2020 after cyclone Iota and Eta to the governments of Nicaragua (CCRIF 2021b). However, the estimated losses in those cases were $3.4 billion for Bahamas (Deopersad et al. 2020) and $743 million for Nicaragua (Reuters 2020). In both cases, the payouts only contributed to 0.4% and 4.1% of the estimated losses, respectively. Therefore, the governments should not just plan for hedging the flood risk through insurance, reinsurance, and CAT bonds but they should also plan for improving flood resilience so that the damage can be limited to its minimum.

Flood risk reduction by single structures such as dams, levees, etc., has been investigated extensively by previous researchers. However, few of them have considered a combination of multiple strategies in flood risk reduction. van Berchum et al. (2019) have developed multiple lines of defense optimization strategies (MODOS) model to plan for multiple lines of flood risk reduction while considering the interdependencies between the strategies. The article considers four types of strategies. They are flood defenses such as levee, storm surge barrier, nature-based solutions such as wetland soyster reefs, damage restricting measures such as flood proofing buildings by slab elevations, and evacuation of flood vulnerable zones.

Infrastructures are necessary to reduce vulnerability and improve resilience of any community, city, or country (ISDR 2005). Infrastructures not only protect against natural hazards, but they are also essential for economic development and reducing poverty (UN 2011). In recent years, the local governments and public works officials have expressed increasing interest in green infrastructures to manage and mitigate flood risk (Carter et al. 2019). O’Donnell et al. (2020) has argued that finding an optimal balance between grey and blue and green infrastructures for maximizing flood risk reduction is the key to the planning of successful flood resilient cities. Investment in improving flood resilience can have long term benefits. For instance, a report published in 2019 by the U.S. National Institute of Building Sciences noted that every dollar invested in adopting the latest building codes can save six dollars from riverine flood damage yielding a benefit to cost ratio of 6:1 (Multi-Hazard Mitigation Council, 2019). The United Nations Environment Program (UNEP) has developed a tool RiVAMP (Chatenoux et al. 2012) to quantify the role of ecosystem in disaster risk reduction.

However, Kundzewicz et al. (2018) have claimed that structural measures such as dams and levees create an illusion of false safety among the populations protected by those dams and levees. Thus, various nonstructural measures should also be considered. Non-structural measures can be insurance, disaster management, early warning system, etc. Hegger et al. (2014) have suggested developing a diversified portfolio of flood risk reduction strategies that combine flood risk mitigation, adaptation, and recovery strategies to maximize the benefits from them.

A recent report from Deloitte (Bachir et al. 2019) has provided five recommendations for the insurance industry to navigate through the climate risk landscape. The report recommends that insurance companies should incentivize policy holders who invest in mitigating climate risk by reducing insurance premiums or assisting them in financing the risk mitigation measures. This can help the insurance companies in containing the claims through these adaptation measures. Similar suggestions on insurance companies helping organizations in mitigating climate can be found in another report by McKinsey and Company (Grimaldi et al. 2020). Hudson et al. (2019) have investigated the affordability of risk-based flood insurance and risk reduction measures in Europe. They have found that in absence of risk reduction initiatives undertaken at the households' level, the insurance premiums are likely to double between 2015 and 2055. Therefore, successful flood insurance mechanisms should incentivize risk reduction by the policyholders. Nofal and van de Lindt (2020) have investigated the impact of building level flood risk reduction initiatives such as flood barrier systems, water pumps, etc. Lim and Skidmore (2019) have found that local governments' spending on public safety and welfare significantly reduces flood vulnerabilities.

Property buyout is a popular flood risk reduction choice often adopted for long term flood risk reduction. The U.S. federal government has an existing floodplain property buyout program (CRS 2022). A floodplain buyout refers to a process in which a government agency acquires private property by purchasing it, relocating, or demolishing any existing structures, and preserving the land as open space indefinitely to restore and protect the natural functions of the floodplain. The responsibility for maintaining the acquired parcels of land lies with the local government, and buyout programs typically do not provide funding for future design, maintenance, or utilization of the bought-out land (CRS 2022).

When it comes to property acquisition and demolition, often facilitated by federal funding, a local or state government procures land and structures located in flood-prone areas from willing sellers and proceeds to demolish the structures. Alternatively, state, or local governments acquire land from willing sellers and provide assistance to property owners in relocating to a different site. If the new location falls within a flood zone, the newly constructed structure must adhere to the community's building codes, including requirements such as a certain elevation above ground level. In both scenarios, the land acquired through the buyout process must be preserved and managed as open space.

North Carolina has established a commendable history of incorporating resilience initiatives into its state-level hazard mitigation planning that includes the implementation and maintenance of successful buyout programs (North Carolina Department of Public Safety 2018). Over the period spanning from the mid-1990s to 2019, it is estimated that North Carolina acquired over 5600 homes that were either damaged by hurricanes or situated in flood-prone areas based on their location. These endeavors yielded further benefits when the state achieved FEMA's Enhanced Hazard Mitigation status in 2014, which granted the state access to an additional 5% in funding from the Hazard Mitigation Grant Program (HMGP) after Hurricane Matthew, resulting in an extra $25 million allocated towards assisting 210 homeowners in their relocation efforts (North Carolina Department of Public Safety 2018).

In Iowa, which is one of the severe flood prone states in the U.S., the federal buyout programs have acquired nearly 3000 properties between 2007 and 2017 (Yildirim and Demir 2021). Yildirim and Demir (2021) have developed a web-based environment to analyze property level and community level benefit-cost analysis for property buyouts. Johnson et al. (2020) have found that the cumulative loss from flood damage will exceed the cost of land acquisition for one-third of the unprotected natural land in the U.S. by 2070. They have therefore argued for flood risk reduction through land acquisition and conservation of natural land in the floodplains.

Although these flood risk reduction strategies have been assessed and analyzed for quite some time, lack of funding remains one of the key challenges in risk reduction and resilience planning (Thomalla and Larsen 2010, Malalgoda et al. 2014, Ludin and Arbon 2017). Infrastructure in the U.S. suffers from continuing underinvestment. The most recent report from the American Society of Civil Engineers (ASCE) in 2021 reported that the estimated gap in infrastructure investment could reach nearly $2.59 trillion by 2029 and this investment gap can potentially cost every American household approximately $3300 every year (ASCE 2021). Under this situation various state and local governments introduce innovative funding mechanisms to finance new infrastructures. For instance, the Department of Watershed Management in Atlanta, Georgia issued Environmental Impact Bond (EIB) for financing green infrastructure projects. The EIB was structured in a way to allow the issuer to pay a performance bonus to the investors if the benefits from the green infrastructure projects exceeded the expectations (Hallauer et al., 2019).

Redirecting of post-disaster aids to pre-disaster investments appears to be the largest opportunity for financing flood resilience (Colgan et al., 2017). Reguero et al. (2020) have explored how the benefit of resilience building projects like reef restoration can be realized in terms of reduction of flood insurance premium due to reduction of flood risk as an outcome of the resilience building projects. Therefore, they have argued that the derived benefit in terms of reduced flood insurance premium should be invested in resilience building. Vaijhala and Rhodes (2015) have proposed a concept of resilience bond. Resilience bonds are a special type of CAT bonds that links the insurance premium with resilience building projects. A part of the bond proceeds is invested in resilience. As an outcome of this investment, the insurance premium is reduced for the policy holder. The investors of the resilience bonds also benefit from the reduced risk to their principal (Vaijhala and Rhodes 2018). Regarding flood risk reduction and resilience, it has been established that an individual organization or stakeholder cannot successfully understand and resolve flood risk (APSC 2012). Therefore, the possibility of engaging flood related insurance, reinsurance sectors in addition to the CAT bonds in flood resilience planning and financing should not be neglected.

2.4 National Flood Insurance Program

In U.S., the National Flood Insurance Program (NFIP) started in 1968 under National Flood Insurance Act. The reluctance of private insurers to offer flood insurance created the need for the

NFIP (Kousky et al. 2020). The NFIP offers flood insurance to individuals, businesses, and renters, providing them with vital protection and expediting their recovery process once floodwaters recede. Collaborating with communities, the NFIP mandates the implementation and enforcement of floodplain management regulations, effectively mitigating the impact of flooding. Flood insurance is accessible to residents and businesses within nearly 23000 NFIP-participating communities. Presently, in communities participating in the regular program, the highest coverage limit for residential buildings with one to four units is set at $250,000, while coverage for personal belongings is capped at $100,000. For residential buildings categorized as Other Residential, comprising five or more units (excluding condominiums), the maximum coverage limit is $500,000 for the structure itself, accompanied by a contents maximum limit of $100,000 (FEMA 2021).

Historically, the private flood insurance market in the U.S. has been relatively small. It typically offers coverage beyond the NFIP limit or in cases where lenders enforce insurance on non-compliant properties (Dixon et al. 2007). While there has been some recent growth in the private sector, the number of private policies remains significantly lower than the NFIP policies. Moreover, private flood insurance often targets high-value homes with higher premiums, often combining flood and homeowners' coverage and offering better protection (NFIP 2015).

The NFIP has undergone multiple revisions since its establishment. These revisions were aimed to secure its financial stability and enhance its mapping and rate-setting processes (FEMA 2021). Purchasing flood insurance was voluntary till 1973. Since then, buying flood insurance was mandated for properties with mortgage from a federally regulated or backed lender which are in a NFIP participating community and within 100-year flood zone by the Flood Disaster Protection Act of 1973 (FDPA). The National Flood Insurance Reform Act of 1994 bolstered the NFIP through various reforms aimed at enhancing lender compliance, introducing mitigation insurance, and establishing a mitigation assistance program. These measures were designed to effectively mitigate the costly and devastating consequences of flooding. The Flood Insurance Reform Act of 2004 reinforced the NFIP by implementing several reforms. These reforms aimed to minimize losses for properties that had repetitive flood insurance claims, raise awareness among policyholders regarding their individual flood insurance policies, provide comprehensive information to policyholders about the claims process, and introduce a minimum criteria for training and education for insurance professionals in flood insurance. The Biggert-Waters Flood

Insurance Reform Act of 2012, known as Biggert Waters, approved and allocated funds for the national mapping program and implemented rate adjustments to ensure the financial stability of the program. These adjustments aimed to transition the program from artificially low subsidized rates to full actuarial rates that accurately reflect the associated risk. The Consolidated Appropriations Act of 2014 included provisions that halted the enforcement of specific sections of the Biggert-Waters law, which were related to rate increases. This action was taken to address concerns regarding the potential impact of rate increases, while new legislation was being developed to address these concerns. The Homeowner Flood Insurance Affordability Act of 2014 (HFIAA) revoked specific provisions of the previous law, Biggert-Waters, that resulted in the reinstatement of grandfathering, the imposition of restrictions on certain rate increases, and the adoption of a revised approach to maintaining the financial stability of the fund through the implementation of an annual surcharge for all policyholders.

The premium for NFIP, i.e., the rating structure has remained largely unchanged since the 1970s. It was based on conventional insurance practices at the time of its establishment. Properties were categorized and assigned rates based on key factors including their location within a flood zone as indicated on a Flood Insurance Rate Map (FIRM), the type of occupancy, and the elevation in relation to the Base Flood Elevation (BFE). On the Flood Insurance Rate Map, flood hazard areas are designated as Special Flood Hazard Areas (SFHAs), which represent the regions that are projected to be inundated by a flood event with a 1-percent chance of occurring or exceeding in any given year. This 1-percent annual chance flood is commonly known as the base flood or 100-year flood. SFHAs are categorized and labeled on the map as Zone A, Zone AO, Zone AH, Zones A1-A30, Zone AE, Zone A99, Zone AR, Zone AR/AE, Zone AR/AO, Zone AR/A1-A30, Zone AR/A, Zone V, Zone VE, and Zones V1-V30. Additionally, the FIRM also indicates moderate flood hazard areas, identified as Zone B or Zone X, which lie between the boundaries of the base flood and the 500-year flood zones. Areas outside the SFHA that are situated at elevations higher than the 0.2-percent-annual-chance flood are classified as Zones C or Zone X, representing minimal flood hazard areas. However, this rating system did not consider the specific flood risk of individual properties, or the costs associated with rebuilding, and it only accounted for two sources of flood risk: river flooding and coastal flooding (CRS 2023).

As a federally run program, NFIP premium costs are highly subsidized, especially for buildings that were constructed before FEMA adopted Flood Insurance Rate Maps, known as Pre-

FIRM buildings. These buildings are constructed on or before December 31, 1974 (FEMA 2021). Although many Pre-FIRM buildings are exposed to severe flood hazard, their NFIP premiums are subsidized due to concerns that those homeowners did not expect high flood insurance cost when their homes were constructed (Miller et al. 2019). The subsidized NFIP premiums have raised some concerns among the researchers. For instance, Young (2008) has claimed that subsidies are not anticipated to be derived from a wide range of taxpayers. Instead, they are projected to be generated by requiring participation within the floodplain areas. The research has further claimed that the NFIP requirements have been designed to protect the lenders and not the flood victims. Despite the subsidized rate, the NFIP penetration rate has been quite low (Kousky 2011). According to a study conducted by Chappell et al. (2007) using Gulf Coast Mississippi communities' survey data, it was found that 69 percent of residents did not possess federal flood insurance. FEMA also acknowledged the significant lack of coverage among residents residing in floodplains. In 2006, Butch Kinerney, a spokesperson for NFIP, stated that flood insurance was held by only 44 percent of the homes that were supposed to have it (Boreczky 2006). As explained in the previous chapter and demonstrated in Figure 1.2, historically most of the disaster survivors have been found uninsured or underinsured (Horn 2019). The penetration rate is even lower in the areas that are outside the 100-year flood zone (Kousky 2011).

Due to this low penetration rate, researchers have been trying to understand the factors that contribute to the demand for flood insurance. For instance, Browne and Hoyt (2000) have used historical data between 1983 and 1993 to understand the determinants of flood insurance demand such as mitigation spending, federal assistance, premium price, income, federally backed mortgage, and recent flood event using fixed effects model. They found a positive relationship between income and the extent of flood insurance coverage obtained. Their empirical findings indicate a negative correlation between the cost of flood insurance and number of flood insurance policies purchased, which is in line with the inverse relation between price and demand for a good or service. Furthermore, research offers evidence that the number of flood insurance policies sold in a year is positively impacted by flood losses experienced in the previous year. Lastly, they found that the availability of disaster relief increased the demand for flood insurance.

Kriesel and Landry (2004) have estimated the impact of raising the flood insurance premium on the coastal areas by random utility maximization for insurance purchase decision. The results are similar to that of Browne and Hoyt (2000) as they have also found that price had a negative impact on the demand for flood insurance. On the other hand, higher income households were more likely to have flood insurance. They found that chances of having flood insurance decreases as the location of a home move further away for coastal flood zones. The findings indicate that households situated near artificial shoreline protection have a 12 percent higher likelihood of acquiring flood insurance.

Michel-Kerjan and Kousky (2010) have analyzed the flood insurance market in Florida, which has over 40 percent of the total NFIP policies-in-force in the U.S., for examining the attributes of homeowners who make the decision to purchase flood insurance. They found that 80% of the flood insurance is for single-family residential properties. Unsurprisingly, about 75% of all single-family insurances were in the 100-year flood plain. They reported that the deductible choices differed based on the flood zone, as individuals in the highest-risk areas, subject to the mandatory purchase requirement, opted for a higher deductible. Lastly, consistent with previous findings, they also found evidence of a response to the 2004 floods in Florida, as homeowners opted for lower deductibles and higher coverage limits compared to their previous choices.

Kousky (2011) investigated the demand for flood insurance utilizing data from all flood insurance policies in St. Louis County, Missouri, between 2000 and 2006. The study revealed low take-up rates, with initial policy underwriting being a more significant challenge than policy retention. Take-up rates increased with larger amounts of land located in both 100-year and surprisingly, 500-year floodplains. In regions with levee protection and even along major rivers, there was a decline in take-up rates. Homeowners situated beyond the 100-year floodplains, exempt from the mandatory purchase requirement, chose policies with lower deductibles and broader coverage. The research also revealed a correlation: higher coverage was sought by homeowners with greater property value and median income, and closer proximity to major rivers.

Michel-Kerjan et al. (2012) analyzed the NFIP policy tenure using data between 2001 and 2009. They found that the average duration of new policies within that period ranged from two to four years and remained relatively consistent across different levels of flood risk. The length of policy tenure was influenced by previous flood experiences, as individuals who have filed small flood claims tend to maintain their insurance for a longer period, while those who have faced significant flood claims are more prone to letting their insurance lapse sooner. They suggested that other pressing needs of daily life such as buying healthier food, replacing bald tires, buying health insurance, etc., might force homeowners to prioritize other expenses over buying and continuing flood insurance, especially those who live from paycheck to paycheck.

Petrolia et al. (2013) have found that homeowners who anticipate higher flood damage are more likely to insure themselves from flood damage. Several studies have found that the demand for flood insurance increases after a flood event (Gallagher 2014, Atreya et al. 2015, Ren and Wang 2016), but may not be sustained as the bump dies down after 3 years (Kousky 2018). While there has been a general consensus on the factors that impact the demand for flood insurance, researchers are yet to agree on how the availability of post-disaster federal assistance influences the NFIP take-up rate. It has been previously found that the availability of post-disaster government assistance reduces the demand for flood insurance, an event popularly known as Charity Hazard (Browne and Hoyt 2000).

Charity hazard can be considered as a specific instance of the moral hazard issue, which refers to a change in risk-taking behavior due to an insurance contract, as individuals may take more risks when they have insurance. This phenomenon occurs when there is a lack of symmetric information between the insurer and the policyholders, preventing the insurer from observing changes in risk-behavior and adjusting premiums accordingly. Government-provided financial assistance acts as insurance against natural hazards without any premium charged to citizens. Consequently, the incentive to buy insurance is diminished. Similar to moral hazard, charity hazard arises when the expectation of governmental relief leads to behavioral changes, causing individuals to forgo insurance and preventive measures. However, charity hazard does not rely on asymmetric information. Furthermore, the implications for insurance companies differ, as charity hazard does not impact the insurer's solvency.

There is a plethora of research on this topic. Raschky et al. (2013) have analyzed the issue of charity hazard for flood insurance in Austria and Germany based on homeowners' willingness to pay (WTP) for flood insurance. Through a Tobit regression model, they found that the expected government relief program had a strong crowding out effect on the demand for flood insurance in both Austria and Germany when income, flood damage, household size, and other factors are controlled. They have also found that the effect is stronger when the availability of this relief is certain. However, their research considered the expected government relief as a binary variable with possible values, i.e., yes (if available) or no (if not available).

Ren and Wang (2016) have analyzed the occurrence of charity hazard in a Chinese context. They found Chinese homeowners' willingness to buy (WTB) and willingness to pay (WTP) is negatively influenced by the relief they received from governments and charity. Thus, they confirm the existence of charity hazard in the Chinese flood insurance market. Through laboratory experiments and surveys, researchers have investigated whether individuals take disaster aid into account when making insurance choices. While these studies provide valuable insights into how pre-disaster perceptions can influence purchase decisions, survey responses may not always align with real-world behavior. Findings from these studies vary. Typically, when individuals are informed about available assistance, it does lead to a decrease in their willingness to pay for insurance (Kousky 2018). However, in the absence of such prompts, disaster aid might not factor into their insurance decision-making process at all. (Botzen and van den Bergh 2012, Petrolia et al. 2013, Raschky et al. 2013).

Andor et al. (2020) have also analyzed the occurrence of charity hazard in the German flood insurance market. They conducted a differentiated analysis of charity hazard, analyzing the impact of government assistance on households in various flood-prone areas, taking into account diverse precautionary measures. Their results demonstrate sufficient heterogeneity in the effects of governmental relief. Trust in relief from the government was positively correlated with the implementation of structural protection initiatives. However, for insurance demand, the relationship varies depending on flood exposure. People in flood exposed areas who trust in government aid are less likely to purchase flood insurance. Conversely, for those with less flood exposure, there is no robust relationship between trust in governmental relief and insurance demand. Their findings suggest the existence of charity hazard in flood insurance uptake for households in flood-exposed areas.

In Tesselaar et al. (2022), a partial equilibrium model was used to examine charity hazard and the insurance gap in European Union countries till 2050. The analysis employed the expected utility framework with decision functions for insurance buying that accounted for the likelihood, uncertainty, and extent of government assistance. By considering country-level insurance systems, government assistance types, and flood risk, the research evaluated the development of charity hazard under varying conditions. Findings indicated that charity hazard diminished with greater uncertainty regarding government compensation and higher flood risk.

Although the evidence has been strongly in favor of the existence of charity hazard in Europe and China, the results are contradictory in the U.S. As explained earlier, Browne and Hoyt (2000) have found that the post-disaster federal assistance increased the demand for the NFIP. They have used historical data between 1983 and 1993 in a fixed effects regression model to explain the demand for NFIP in terms of disaster assistance. The results show that the regression coefficient is positive and statistically significant at 5% significance level. Therefore, an increase in the disaster assistance is expected to increase the demand for flood insurance in the U.S., which refutes the existence of charity hazard. However, their study did not consider the lag between disaster relief and flood insurance enrollment. The expected effect of disaster relief may not be apparent immediately as it might take some time to purchase flood insurance after a disaster. Therefore, there is a necessity to consider the time lag between disaster relief and flood insurance enrollment.

Petrolia et al. (2013) have conducted a household survey in the U.S. Gulf coast states to understand the factors that impact the decision of a household to purchase flood insurance. The research employed a Probit regression model that expressed the binary decision of purchasing or not purchasing flood insurance based on different variables such as risk perception, risk aversion, disaster assistance, insurer credibility, past experience with floods, etc. Disaster assistance was considered a binary variable that measured the expectations regarding receiving post-disaster federal assistance. The regression coefficient for disaster assistance was positive, which indicates that positive expectations regarding the availability of post-disaster federal assistance increased the prospect of purchasing flood insurance. However, as explained earlier, the outcomes of survey-based methods can differ from real-world behavior (Kousky 2018).

Kousky et al. (2018) have examined charity hazard in the U.S. flood insurance market conducting the analysis at the ZIP, i.e., postal code level while considering the time lag between federal assistance (IHP) and flood insurance enrollment. They have developed a panel dataset of the ZIP codes by utilizing historical data between 2000 and 2011. They found that the charity hazard existed for insured amount although it did not exist for take-up rate. However, their analysis only considered the availability of federal assistance as a binary variable (IHP assistance was either available or not available). But that is only one part of the picture. Not all ZIP codes received an equal amount of IHP assistance. In fact, the level of IHP assistance is different in different ZIP codes, which makes the variable continuous. The analysis of the effect of continuous treatment is missing in Kousky et al. (2018).

Davlasheridze and Miao (2019) have conducted the analysis at the county level using the data between 1998 and 2010 to assess the existence of charity hazard in the U.S. Although their analysis is focused on analyzing the impact of Public Assistance (PA), which compensates the State, Local, Tribal, and Territorial (SLTT) governments to repair their disaster damaged infrastructure, on the flood insurance enrollment, they have found that Individual Assistance (IA) increased the flood insurance enrollment. They have developed a fixed effects regression model that expressed per capita take-up rate, policy coverage, number of policies, and total premiums based on different predictors such as rainfall, income, population, and variables that were related to the political affiliation of the county. It is worth mentioning that PA does not compensate households thus is not directly relevant to flood insurance. Additionally, they have considered all the NFIP participating counties regardless of the Presidential declaration of major disaster. Lastly, the annual rainfall amount does not provide the actual extent of flood damage, which influences the demand for flood insurance (Browne and Hoyt 2000, Petrolia et al. 2013).

More recently, Landry et al. (2021) utilized household level survey data to find and quantify expectations regarding eligibility for government aid following a disaster declaration. This approach served as a direct means to examine the decision of households to forego or reduce their flood insurance coverage. Their bivariate probit model results show that charity hazard existed in the U.S. flood insurance market. They have estimated that individuals who had optimistic expectations of qualifying for disaster assistance for home repair were 32.9% less inclined to have a flood insurance policy while controlling for other significant factors such as residing in flood zone, etc.

As explained, previous researchers have found conflicting evidence on the existence of charity hazard in the U.S. flood insurance market. Theoretically, charity hazard is expected to exist as post-disaster government assistance reduces the incentive for households to insure themselves. However, the situation is somehow different in the U.S., where the federal assistance is only made available in the event of a Presidentially declared major disaster. The federal government compensates the households and businesses for the losses from a natural hazard through Individual Assistance (IA) program. The U.S. Housing and Urban Department (HUD) also has two primary sources of disaster relief (Kousky and Shabman 2012). The mortgage assistance program provides insurance to homeowners after a major disaster so that they can get qualified for loans to rebuild their houses. Evidently, the insurance is not free of cost. The other program managed by HUD is Community Development Block Grant (CDBG) which is a non-disaster program but has been used as a disaster assistance to fulfill the needs unmet by other federal assistance programs (Kousky and Shabman 2012). CDBG funds can be used for housing, economic development, clean up, infrastructure recovery, hazard mitigation initiatives after a major disaster. The U.S. Small Business Administration (SBA) also provides low interests loans to households and businesses to expedite their recovery from a presidentially declared major disasters. However, the Individual and Households Program (IHP) within IA is the primary way the US Federal Emergency Management Agency (FEMA) supports disaster survivors (Webster 2019).

The individuals and households program (IHP) provides direct financial assistance to qualified individuals and households who are underinsured or uninsured and have serious needs of support as a result of a Presidentially declared emergency or major disaster. Disbursement of IHP in case of a national emergency is rare. It is primarily used for major disasters. The U.S. president can declare a major disaster for any natural hazard such as flood, severe storm, hurricane, earthquake, wildfire, etc., if he or she determines that the damage has exceeded the capability of the state and local government to respond effectively. By declaring a major disaster, federal resources are made available to the state and local governments to help them overcome the disaster. The state or local government does not automatically receive assistance from the federal government. Instead, state or local government must formally petition the president to declare a major disaster, thereby enabling the allocation of resources.

It should be noted that IHP is the only federal assistance program that provides direct financial assistance to disaster survivors. To be eligible to receive IHP assistance, the applicant must furnish that (1) the damage is not uninsured, (2) he or she is a citizen (or qualified alien), and (3) the property is the primary residence (Kousky and Shabman 2012). Importantly, if the applicant resides in an area designated as a 100-year flood zone and in a community that is not enrolled in the NFIP, IHP assistance cannot be utilized for flood-related repairs. This policy is implemented to incentivize communities to join the NFIP (Kousky and Shabman 2012).

It should be noted that the IHP program is purposed to fulfill the basic needs and it does not compensate for all losses (Webster 2019). The program provides Housing Assistance that covers lodging expense, home repair, temporary housing units, etc., and other needs assistance

(ONA) such as childcare assistance, critical needs assistance, transportation assistance, etc. For housing assistance, the federal government bears 100% of the cost whereas for the ONA, the cost is shared between the federal government and the state or local government. Typically, the federal government pays for 75% of the cost and the non-federal share is 25%. Figure 2.4 shows the types of housing and ONA provided by FEMA through the IHP program. The maximum assistance for housing related needs and ONA provided by FEMA through the IHP program is $35500 (Webster 2019). It should be noted that the federal government pays for 75% of the ONA while the rest is paid by the state governments (Kousky and Shabman 2012).

Housing Assistance: Financial	Housing Assistance: Direct	ONA: SBA-Dependent	ONA: Non-SBA-Dependent
Lodging Expense Reimbursement Rental Assistance Home Repair Assistance Home Replacement Assistance	Multifamily Lease and Repair Transportable Temporary Housing Units Direct Lease Permanent Housing Construction	Personal Property Moving and Storage Transportation Assistance Group Flood Insurance Policy	Funeral Assistance Medical and Dental Assistance Childcare Assistance Assistance for Miscellaneous Items Critical Needs Assistance Clean and Removal Assistance

Figure 2.4 Types of Housing Assistance and Other Needs Assistance (Source: Webster 2019)

For homeowners who live in 100-year flood zone and in a community that participates in NFIP, having flood insurance is a requirement for receiving the IHP aid. In absence, FEMA may purchase the flood insurance through the Other Needs Assistance (ONA) funds for 3 years. At its expiration, to qualify for future assistance, the applicant must procure and sustain flood insurance coverage (Webster 2019). Kousky (2013) has investigated the disbursement of IA after major floods, storms, and tornados in Missouri in 2008. She found that the majority of the aid grants were too small, on the order of a few thousand dollars. The average IA grant was approximately $2000. Also, more than 50% of the applicants were not granted aid as they were either ineligible or the damage was considered insufficient. Among different types of housing assistance, as shown in Figure 2.4, most of the IHP was related to home repair, followed by rental assistance, replacement assistance, etc. Among the other needs assistance, most applicants received IHP for personal property damage. The paper also highlighted the difficulties in estimating the extent of damage accurately after the disaster.

The inadequacy of the IA aid has also been reflected in Sterett (2015). Wu et al. (2017) have found that rate of approval for housing assistance through IHP was less than 50% (44% for owners and 32% for renters) after hurricane Harvey in 2017. Hooks and Miller (2006) have claimed that IHP encompassed procedural hurdles that are biased against lower income population thus making it tougher for them to cope. Grube et al. (2018) have concluded that the procedural complexities may have disadvantaged the less educated population based on empirical evidence from superstorm Sandy in 2012. Emrich et al. (2020) have also found that IHP assistance is not systematically linked to social vulnerabilities. Drakes et al. (2021) have found co-occurrence of low level of IHP assistance and high social vulnerability in rural areas and that in Southeastern contiguous U.S. Bann (2020) has analyzed the correlation between individual and community level demographic variables and the outcomes of the IA program. The research has found that variables like homeownership, homeowners' insurance status, income, households' composition are significant in predicting the total assistance amount.

Although the take-up rate for NFIP is quite low, the program is still not profitable. The NFIP has been considered problematic almost since its inception. Schilling et al. (1987) found that the program had been largely unsuccessful in the coastal areas due to paying more on claims than the collected premiums. This problem can be partially attributed to FEMA's inability to charge premiums based on true flood risk. The inability originates from technical as well as bureaucratic challenges. As explained earlier, NFIP has been using flood maps for determining the flood insurance premiums for households and businesses. The U.S. Department of Homeland Security (DHS) reported in 2017 that only 42% of those flood maps can adequately identify flood risk, i.e., majority of them were outdated and could not reflect the true flood risk of a property (DHS 2017). Again, 1st Street Foundation found that the number of properties that are located within the 100-year flood zone is 1.7 times of what the FEMA has estimated (1st Street Foundation 2020). Due to this discrepancy in the flood maps, it has been claimed on several occasions that the NFIP premiums do not reflect the true flood risk of the property that is being insured (Kousky 2018). On the other hand, the Homeowner Flood Insurance Affordability Act of 2014 (HFIAA) has restricted FEMA's ability to increase the premium more than 18% annually.

Payment of flood insurance claims is one of the key operating expenses of the NFIP. Historical records showed that between 1978 and 2017, NFIP collected $60 billion in premiums while paid $65 billion as payouts (Grigg 2019). Following the devastation caused by Hurricane Katrina in 2005, the NFIP has been burdened with a significant debt of $20.5 billion owed to the U.S. Treasury. In 2017, the NFIP program reached its debt limit of $30.4 billion, despite federal taxpayers already shouldering $16 billions of NFIP's debt. These debts represent a liability that is ultimately borne by federal taxpayers. In 2022, the program was projected to pay more than $280 million in interest on this debt. However, raising NFIP premiums has posed significant political challenges due to the potential adverse effects on housing affordability (Miller et al. 2019). The program also pays one-third of its income from the collected premiums to the financial intermediaries for underwriting the flood insurance policies although none of the flood risk is borne by these intermediaries (Grigg 2019). Repetitive loss properties account for 25-30% of the claims although they are only about 1% of the insured properties (Grigg 2019).

Researchers have long been recommending reforms in NFIP. Michel-Kerjan and Kunreuther (2011), Akabas (2014), McShane and Wie (2019) have recommended risk-based premiums, unsubsidized rates, protection of low-income groups, forgiveness of the debt to the U.S. Treasury, reduction of exposure by reinsurance and CAT bonds, etc., for reforming the NFIP. However, affordability of the NFIP policies remains one of the key challenges for the lower-income households (Shively 2017). Frazier et al. (2020) claimed that an NFIP reform without any considerations for the socioeconomic vulnerability will create barriers for lower-income residents. On the other hand, Wagner (2022) has found that the willingness to pay for flood insurance is remarkably low in the U.S. The NFIP's ability to pay off its debt is hindered by recurrent and costly flooding incidents. Currently, the program relies solely on premiums to cover the interest on previous losses, necessitating a shift in this approach to enhance its financial stability. The NFIP must implement a robust financial framework that strikes a balance between affordability and fiscal soundness to ensure its long-term viability and sustainability. To achieve this goal, FEMA has proposed 17 legislative reforms and actions to Congress in May 2022 for consideration during the NFIP's reauthorization process. In a letter to the Congressional leaders, FEMA proposed (1) to make the NFIP more affordable to low-and-moderate income households to increase the flood insurance demand, (2) to build resilience by improving risk communications and providing households with tools to manage flood risk, (3) strengthen local floodplain management standards and address extreme repetitive loss properties, and (4) institute a sound financial framework that balances affordability and fiscal soundness.

In 2021, FEMA also brought a new approach named Risk Rating 2.0 that calculates the flood insurance premium of a household or business based on its true flood exposure thus making it more accurate. All new policies from October 1, 2021, onwards are subject to this new risk rating method. FEMA sees this new approach as a transformational leap forward that will set flood insurance premium rates fairer and more equitable. FEMA estimated that Risk Rating 2.0 would immediately decrease the monthly flood insurance premium for 23% of the policy holders. While 66% of them will see their monthly flood insurance premium increase by less than $10. The remaining 11% will face a monthly increase of flood insurance premium by more than $10 (FEMA 2022). However, there has been some controversy regarding this Risk Rating 2.0, as it is predicted that some states in the U.S. such as Louisiana could see an increase in premium for more than 80% of the existing policies (Murphy 2022).

It has been explained earlier that NFIP's long term solvency and financial issues cannot be simply solved by raising the flood insurance premiums as it might further reduce the demand for flood insurance. The reduced demand could increase the extent of uninsured losses from floods in future. On the other hand, it is expected that the frequency and severity of natural hazards would increase in the long term due to climate change (Smith 2023). Therefore, it is essential that NFIP plans for flood risk reduction to keep the program financially viable. Other than providing flood insurance to households, NFIP has a long-term objective to reduce federal expenditures on post-disaster assistance (Horn and Webel 2021). Therefore, NFIP is expected to ensure that future payouts are kept within limits. This requires the understanding of the causes that influence the annual NFIP payout, i.e., a causal model that explains the flood insurance payouts based on different flood risk factors so that appropriate flood risk reduction strategies and/or policies can be planned to mitigate the impact of those flood risk factors on the NFIP payouts.

2.5 Flood Risk Factors

Risk factors are common in clinical science and defined as the factors that increase the likelihood of developing a disease. Similarly, this research has defined flood risk factors as the factors that increase the likelihood of flood losses and subsequently the flood risk in a region. Since flood risk factors influence the extent of flood loss, it is safe to assume that they influence the extent of flood insurance claims due to flood loss. Identifying the flood risk factors is an essential step for effective flood risk management (Koc and Işık 2021). Table 2.3 provides a list of flood risk factors that have been previously considered in flood loss and flood risk related literature. Yang et al. (2013) have considered 4 categories of flood risk factors that are hazard factors, location characteristics such as vegetation coverage, drainage density, etc., property characteristics such as population density, agriculture, etc., and societal bearing capabilities in terms of early warning systems, disaster relief agencies, etc. Koc and Işık (2021) have clustered flood risk factors into 14 categories that include weather conditions, environmental factors, basin and flood characteristics, institutional capacity, existing condition of infrastructure, land use pattern, demographic and social factors, health, economy, ecology, and accessibility.

Pathak et al. (2020) have found 31 factors that influence the flood vulnerability in Nepal. Most of the factors are related to the socio-economic characteristics of the households. Balica et al. (2012) introduced a flood vulnerability index for cities, considering exposure, susceptibility, and resilience factors. They applied a system-based methodology, dissecting flood vulnerability into hydro-geological, socio-economic, and politico-administrative components. Within these, they identified 19 indicators encompassing aspects like sea level rise, storm surges, cyclone frequency, and population density near coastlines. Terti et al. (2015) outlined dynamic flood vulnerability components: exposure, sensitivity, and coping capacity. Yang et al. (2018) structured their flood vulnerability index around exposure, sensitivity, and adaptive capacity, using indicators such as flood velocity, water depth, and early warning capabilities. Karagiorgos et al. (2016) delineated vulnerability into physical and social components. For physical vulnerability, they established an empirical relationship correlating loss degree with flash flood intensity. This degree of loss was determined by the ratio of collected loss to property value, with the resulting distribution used to quantify physical vulnerability.

Table 2.3 List of Flood Risk Factor

Flood Risk Factors	Description	Supporting Literatures
Precipitation	Average rainfall	Tarhule (2005), Zhang et al. (2018)
Annual Frequency	Annual frequency of floods and related natural hazards	OLOGUNORISA (2004), Gayen et al. (2022)
Historic Severity	Severity of past floods	Botzen et al. (2009), Boamah et al. (2015)
Area	Area of the geographic location	Ouma and Tateishi (2014), Moreira et al. (2021)
Population	Population of the geographic location	Rasch (2016), Moreira et al. (2021)
Age of Buildings	Average age of the buildings	Penning-Rowsell et al. (2005), Koc and Işık (2021)
Building Value	Dollar value of the buildings	Schröter et al. (2014), Wing et al. (2020)
Agricultural Value	Dollar value of agricultural output	Yang et al. (2013), Yang et al. (2018)
Race	Percentage of Hispanic, White, Black, and American Indian and Alaska Native population	Remo et al. (2016), Tate et al. (2021)
Elderly Population	Percentage of Population over 65 years old	Botzen et al. (2009), Dandapat and Panda (2017)
Gender	Percentage of female population	Pathak et al. (2020), Tate et al. (2021)
Disability	Percentage population with disabilities	Campbell et al. (2020), Moreira et al. (2021)
Income	Gini-index of income inequality, Below poverty level population percentage	Dandapat and Panda (2017), Tate et al. (2021)
Education	Percentage population with a high school degree	Botzen et al. (2009), Ahmad et al. (2016), Tate et al. (2021)
Workforce Participation	Percentage population participating in workforce	Kuhlicke et al. (2011), Tate et al. (2021)
Mobile Homes	Number of mobile homes	Baker et al. 2014, Rumbach et al. 2020

Table 2.3 continued

Flood Exposed Buildings	Dollar value of the buildings that are exposed to flood and flood related hazards.	Stephenson and D'ayala (2014), FEMA (2021)
Flood Exposed Population	Number of people exposed to flood and flood related hazards	FEMA (2021), Moreira et al. (2021)
Historic Loss Ratio for Buildings	Percentage of the buildings that have experienced losses in the past.	FEMA (2021), Vishnu et al. (2021)
Historic Loss Ratio for Population	Percentage of the population that have experienced losses in the past.	Schröter et al. (2014), FEMA (2021)
High and Significant Hazard Dams	Number of dams that have been classified as significant and high hazard by the National Inventory of Dams (NID).	Yerramilli (2013), Day (2016)
Infrastructures Vulnerability	Susceptibility of infrastructure to get damaged during floods	Len et al. (2018), Sanders et al. (2020)
Condition of Existing Infrastructures	Existing condition of civil infrastructure	Tu et al. (2011), Deria et al. (2020), Porter et al. (2021)
Capacity of Civic Infrastructures	Institutional or governance capacity	Cutter et al. (2014), Choi et al. (2019)
Community Capital	Community capital	Cutter et al. (2014), Choi et al. (2019)
Flood Insurance	Number of flood insurance policies	Owusu-Ansah et al. (2019), Moreira et al. (2021)
Total Insured Value	Dollar value of the total policy coverage of all flood insurance policies	Patankar and Patwardhan (2016), Wang and Sebastian (2021)
Environmental Protection	Acres of forest land	Bradshaw et al. (2007), Kim et al. (2019)
Availability of Internet	Percentage of households with a broadband connection	Deria et al. (2020), Rashetnia and Jahanbani (2021)

Miguez and Veról (2017) devised a flood risk index centered on flood attributes (depth and duration) as well as repercussions (dwelling density, income per capita, and inadequate sanitation). This flood risk index was subsequently utilized in constructing a flood resilience index. In a study spanning from 1996 to 2015, Lim and Skidmore (2019) scrutinized historical NFIP data and identified that lower economic and social foundations, reduced educational attainment, and substandard housing quality amplify flood vulnerability. Botzen et al. (2009) conducted a survey involving around 1000 homeowners in the Netherlands. They pinpointed various factors, including previous encounters with flood hazards, age, and education, that influence homeowners' perceived flood risk.

Among those factors, one that has appeared in most of the research is social vulnerability. Social vulnerability includes the socio-economic and demographic factors that increase or reduce the impacts of natural hazards on a community (Tierney et al. 2001, Heinz Center 2002). For instance, Campbell et al. (2020) have found that vulnerable populations suffer the most damage from floods. These people include seniors, people with functional and access needs, people of lower economic status, and other minorities. There are a number of research works that have analyzed the role of multiple socio-demographic factors on flood vulnerability and have found significant impact (Cutter et al. 2003, Zhang and You 2014, Dandapat and Panda 2017, Emrich et al. 2020, Drakes et al. 2021, Koc and Işık 2021).

Researchers have found that the past experience with floods improves the ability to recover from future floods (Boamah et al. 2015). However, Fanta et al. (2019) have concluded that historical memory is not sufficient to protect human settlements from rare catastrophic floods. The age of buildings has been used in the past to model the flood losses to residential buildings (Penning-Rowsell et al. 2005, Wing et al. 2020, Koc and Işık 2021). New buildings often follow imrpoved standards but can be more expensive to repair if damaged. The value of the buildings is another factor that has been used in the past to model flood losses to households (Schröter et al. 2014, Wing et al. 2020). The agricultural value has also been used in flood loss and flood vulnerability assessment (Yang et al. 2013, Yang et al. 2018). It has been found that mobile homes are more prone to flood damage (Baker et al. 2014, Rumbach et al. 2020). The existing coping capacity of a community is an important predictor for estimating the impact of a natural hazard in that community (Scheuer et al. 2011, Yang et al. 2013, Terti et al. 2015). Choi et al. (2019) have proposed that a disaster resilient community needs capacities in all seven layers of infrastructures. These seven layers are civil, civic, social, educational, financial, environmental, and cyber.

2.6 Summary and Point of Departure

The NFIP at its existing condition is under huge debt to the U.S. treasury (Grigg 2019) and the debt is expected to increase in future (CBO 2017) due to increasing frequency and severity of natural hazards. Additionally, the NFIP pays a large sum as interest on that debt. There are several reasons for the insolvency problem. First, the flood insurance take-up rate in the U.S. is quite low (Kousky 2011, Michel-Kerjan et al. 2012, Kousky et al. 2018). Flood insurance is only required for properties that are located in a NFIP participating community and within a 100-year flood zone and have a federally backed mortgage. Recent research shows that NFIP designated 100-year flood zone grossly underestimates flood risk. In fact, 1st Street Foundation found that the number of properties located in 100-year flood zones is 1.7 times higher than what FEMA calculated (1st Street Foundation 2020). Second, the NFIP premiums were not risk based until very recently. The flood maps that were used by FEMA are outdated and the majority of them do not reflect the true flood risk in a location (DHS 2017). FEMA expects to eradicate the premium problem by adopting an actuarial approach named "Risk Rating 2.0", where FEMA claims that the revised premium will reflect the true flood risk of a property. However, due to bureaucratic challenges, FEMA is unable to raise premiums more than 18% annually. Finally, NFIP as a government sponsored program is the insurer of last resort even for the households that are deemed uninsurable by private flood insurers (FEMA 2015, Horn and Webel 2021). As result, it cannot cherry pick households that it wants to insure. This puts additional pressure on the program due to the presence of asymmetrically used information between the insurer and insured (Bradt et al. 2021). Historically, 25% to 30% of the claims have originated from repetitive loss properties, which are only 1% of all the insured properties by the NFIP (Grigg 2019).

As highlighted throughout this chapter, previous studies on this topic such as Michel-Kerjan and Kunreuther (2011), Akabas (2014), Shively (2017), Kousky (2018), Kousky et al. (2011, 2018, 2020), McShane and Wie (2019), Frazier et al. (2020), Wagner (2022), etc., have provided several recommendations for reforming the NFIP. Recommendations such as risk-based premiums, unsubsidized rates, debt cancellation, increased affordability, protection of low-income groups, etc., have appeared regularly in the literature. The majority of these recommendations emphasize reforming the NFIP premiums, which is finally being done through Risk Rating 2.0. However, as explained earlier, this insolvency problem cannot be simply solved by reforming the flood insurance premium as higher premium might further reduce the demand for flood insurance and NFIP might be left only with high-risk properties that would further increase the likelihood of large payouts. On the other hand, decreasing the premium to increase the take-up rate can cause a reduction in revenue, which will put more financial strain on a program that is already under huge debt. Therefore, *what has been missing in the existing literature is a holistic analysis of the problem that produces empirical evidence to help decision makers keep the NFIP running in the long-term*. To fill the identified gap, this research proposes three major tasks, as presented in Figure 2.5, to ensure the long-term sustainability of the NFIP. They are (1) increasing demand for flood insurance, (2) decreasing future payouts through risk reduction, and (3) enhancing financial preparedness by developing the necessary predictive ability.

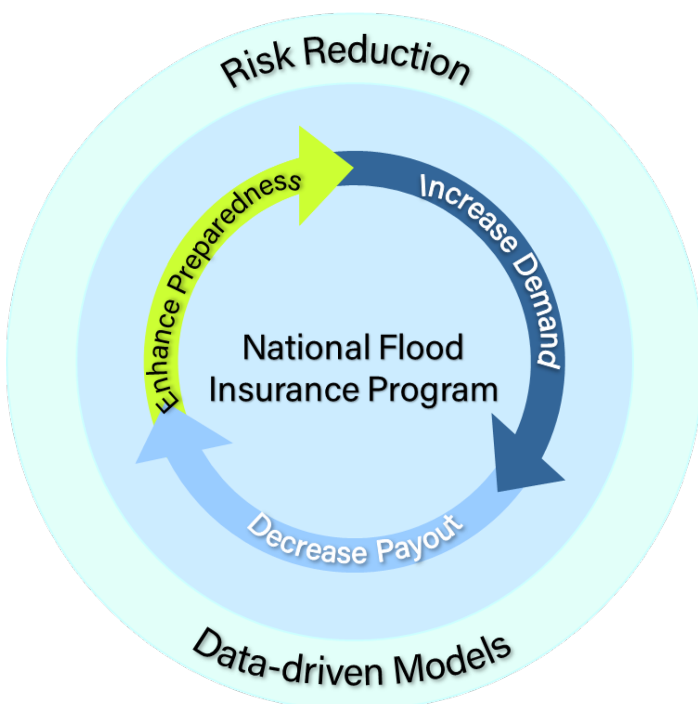


Figure 2.5. Conceptual Framework for Long-Term Sustainability of the NFIP

This research proposes that data-driven empirical insights are essential for designing effective policies that can ensure the long-term sustainability of the NFIP. Additionally, the dissertation hypothesizes that risk reduction is essential in order to keep the future payouts in control. Only through risk reduction, the NFIP can continue to provide flood insurance to millions of Americans and at the same time keep the program solvent. Based on the problem statement and proposed solutions, the research statement has been formulated as follows – to mitigate flood risk through flood insurance, it essential to increase the flood insurance take-up rate and simultaneously decrease the likelihood of large payouts. This can only be achieved through flood risk reduction. Effective flood risk reduction policies can be more efficiently planned if the decision makers have the knowledge of the contributing factors of flood insurance demand and payouts based on data driven robust models.

3. POST-DISASTER FEDERAL ASSISTANCE AND FLOOD INSURANCE DEMAND

Abstract

This chapter tested the hypothesis that the post-disaster federal assistance program reduces the demand for flood insurance in the U.S. To do that, this research has collected data between 2016 and 2020, and conducted propensity score-based analysis to understand the causality between post-disaster federal assistance and flood insurance enrollment in the flood affected counties in the U.S. In the first part, the treatment variable, i.e., the federal assistance has been considered as binary to compare the effect of the availability of federal assistance to that of non-availability of federal assistance on flood insurance enrollment in a flood affected county by using propensity score matching method. It was found that the availability of the federal payout in a county in a year increased the number of flood insurance policies by 5.2% and the total insured value of the policies by 4.6% in the following year. Since the level of federal assistance is different in different counties that received the assistance, in the next step, the treatment variable was considered continuous to estimate the effect of different levels of treatment. Hence, the generalized propensity score method was used to develop a dose response function, i.e., a function that depicts the changes in the outcome variables, i.e., annual NFIP enrollment based on different levels of treatment, i.e., federal assistance. It was found that for each 1000 households in a county that received federal payout, the percentage increase in the number of flood insurance policies was 3.41%. On the other hand, for each million-dollar federal payout in a county, the total insured value of the flood insurance policies increased by 1.96%. Therefore, this research has concluded that contrary to the hypothesis, the availability of post-disaster federal payout increased the flood insurance enrollment in the U.S.

3.1 Introduction

The National Flood Insurance Program (NFIP), which dominates the U.S. flood insurance market, was established by the U.S. federal government in 1968 as a part of the 1968 National Flood Insurance Act (Davlasheridze and Miao 2019). Kousky et al. (2018a) reported that private flood insurance accounted for only 3.5% to 4.5% of all residential flood insurance policies. The primary objective of the NFIP was to offer flood insurance to homeowners and businesses at a subsidized rate to mitigate their loss from floods. Purchasing flood insurance was voluntary till 1973. Since then, buying flood insurance was mandated for properties with mortgage from a federally regulated or backed lender that are located within 100-year flood zone (Insurance Information Institute 2020).

The flood insurance take-up rate in the U.S. is low (Kousky et al. 2018). Hence, the U.S. federal government, as the insurer of the last resort, compensates the disaster survivors who are underinsured and/or uninsured through the Individual Assistance (IA) program. The Individuals and Households Program (IHP) within IA is the primary way the U.S. Federal Emergency Management Agency (FEMA) supports disaster survivors (Webster 2019). IHP provides direct financial assistance to eligible individuals and households who are underinsured or uninsured and have serious needs of support as a result of a presidentially declared emergency or a major disaster. To be eligible to receive IHP assistance, an applicant must prove that (1) the damage is not insured, (2) he or she is a U.S. citizen (or qualified alien), and (3) the property is the primary residence (Kousky and Shabman 2012).

The demand for flood insurance is influenced by several factors such as the premium level, household income, damage from recent flood events, population, post-disaster federal support, etc. (Browne and Hoyt 2000, Kousky et al. 2018, Landry et al. 2021). There is a general consensus about how these factors influence the demand for flood insurance except for post-disaster federal support. It has been found that the availability of post-disaster government support often crowds out the demand for private flood insurance, an event popularly known as Charity Hazard (Browne and Hoyt 2000, Raschky and Weck-Hannemann 2007). In the U.S., the federal regulations require that the IHP recipients maintain flood insurance for future assistance. Therefore, IHP should not crowd out the demand for flood insurance. However, researchers have found conflicting evidence, which have been presented in Table 3.1, on the existence of charity hazard in the U.S. flood insurance market.

This research is a further attempt in answering the question on whether IHP assistance influences the NFIP enrollment in the U.S. while using more recent data, a different approach, and a different method from the previous studies on this topic. In this new approach, the research questions into two parts for deeper insights (1) how the NFIP enrollment differs between the counties that received IHP payout and the counties that did not receive the IHP payout despite the declaration of major flood related disaster thus making them eligible to receive IHP assistance and (2) how different levels of IHP payout influences the NFIP enrollment in a county. To answer these questions, this research has collected data from multiple sources to create an unbalanced panel dataset that has been used in a Propensity Score Matching (PSM) method to answer the first research question and Generalized Propensity Score (GPS) method to answer the second research question. The outcomes of this research will assist the policy makers to decide the future direction of these two federal programs as it can greatly influence the cost of flood events to the federal government. If the IHP crowds out flood insurance, then in future uninsured flood losses will greatly increase as more households will rely on the federal government instead of purchasing flood insurance. On the other hand, this can also increase the risk of large payouts to the NFIP because only the properties with high flood exposure might be left in the pool of NFIP due to presence of asymmetrically used information between insurer and insured (Bradt et al. 2021). Due to the increasing frequency of natural hazards, this domino effect of charity hazard could lead to insolvency of NFIP. Thus, this scenario increases the challenge of running both NFIP and IHP programs. On the other hand, if the IHP reinforces the NFIP enrollment, it will be a win-win for the federal government as more households will purchase flood insurance which will subsequently reduce the cost of running the IHP program. The last scenario is where the impact of the IHP on the NFIP is insignificant and therefore, it neither reinforces nor crowds out the NFIP enrollment. In this scenario, both programs can run independently.

3.2 Research Background

To illustrate the low flood insurance take-up rate, Munich Re reported that only 5% of all the single-family homeowners in the U.S. are insured against flood hazard (Munich Re 2020). They have also reported in 2020 that there were 14.6 million properties in the U.S. that were at substantial flood risk, i.e., located in a 100-year flood zone (Munich Re 2020). On the other hand, NFIP records showed that in 2020, there were approximately 4.03 million flood insurance policy holders. This low take-up rate has also been reported in a congressional research report (Horn 2019). The report records the average NFIP take-up rates across counties in multiple flood events: for instance, South Carolina Flood in 2015 (5%), Louisiana Flood in 2016 (17%), Hurricane Harvey in Texas (10%), Hurricane Irma in Florida (12%). Other studies have found that the take-up rate in 100-year flood zone is high, around 50 percent (Kousky and Michel-Kerjan 2012, Dixon et al. 2013), while very low outside the 100-year flood zone, even if the area is exposed to flood hazard (Kousky 2018).

A number of studies have examined the factors that influence the demand for flood insurance. For instance, Kousky (2011), Landry and Jahan-Parvar (2011) and Ren and Wang (2016) have found that the flood insurance take-up rate is generally higher in the areas of higher flood exposure. They have also found that the homeowners' education level and home value also have positive impact on the demand for flood insurance. Petrolia et al. (2013) have found that homeowners who anticipate higher flood damage are more likely to insure themselves from flood damage. Several studies have found that the flood insurance penetration rate increases following a flood event (Gallagher 2014, Atreya et al. 2015, Ren and Wang 2016), but may not be sustained as the bump dies down after 3 years (Kousky 2018).

Another factor that has been found influencing the flood insurance take-up rate is the availability of post-disaster funding from the governments. As mentioned earlier, charity hazard is defined as the propensity of individuals to not insure themselves against natural hazards because they believed that help would be available from friends, family, charities, emergency agencies, and government (Browne and Hoyt, 2000). This problem results in households under-insuring or not insuring themselves against natural hazards due to the expected government relief. Raschky et al. (2013) have analyzed the issue of charity hazard for flood insurance in Austria and Germany based on homeowners' willingness to pay (WTP) for flood insurance. Through a Tobit regression model, they found that the expected government relief program had a strong crowding out effect on the demand for flood insurance in both Austria and Germany when income, flood damage, household size, and other factors are controlled. They have also found that the effect is stronger when the availability of this relief is certain. Ren and Wang (2016) also found Chinese homeowners' WTP is negatively influenced by the relief they received from governments and charity.

Table 3.1 summarizes the findings from the relevant literature on charity hazard in flood insurance market. It can be noticed that previous researchers have conducted both micro level, i.e., household level and macro level, i.e., ZIP or postal code, or county level analysis to understand the issue of charity hazard in the U.S. flood insurance market. When the analysis is conducted at micro level, the outcomes have predominantly been in favor of the existence of charity hazard. Although, Petrolia et al. (2013) have found that the higher expectations of receiving post-disaster government assistance increases the likelihood of purchasing flood insurance thus rejecting the existence of charity hazard. However, for macro level analysis, the results are conflicting.

Table 3.1 Previous Research on Charity Hazard in the Flood Insurance Market

Literature	Country/Region	Method	Level of Analysis	Existence of Charity Hazard
Raschky et al. (2013)	Austria and Germany	Tobit Model	Micro	Yes
Petrolia et al. (2013)	U.S.	Probit Model	Micro	No
Ren and Wang (2016)	China	Logit Model	Micro	Yes
Andor et al. (2020)	Germany	Probit Model	Micro	Yes
Landry et al. (2021)	U.S.	Instrumental Variable	Micro	Yes
Browne and Hoyt (2000)	U.S.	Fixed Effects Model	Macro	No
Kousky et al. (2018b)	U.S.	Fixed Effects and Instrumental Variable Model	Macro	Yes, for Insured Amount and No, for Take-up Rate
Davlasheridze and Miao (2019)	U.S.	Fixed Effects Model	Macro	No
Tesselaar et al. (2022)	Several Countries in the European Union	Partial Equilibrium Model	Macro	Yes

Among the macro level analysis, Browne and Hoyt (2000) opposed the idea of charity hazard. They found significant positive relation between government disaster relief and demand for flood insurance in the U.S. They conducted their analysis at state level using the data between 1983 and 1993. However, their study did not consider the lag between disaster relief and flood insurance enrollment. The expected effect of disaster relief may not be apparent immediately as it might take some time to purchase flood insurance after a disaster. Therefore, there is a necessity to consider the time lag between disaster relief and flood insurance enrollment. Kousky et al. (2018b) have conducted the analysis at the ZIP, i.e., postal code level while considering the time lag between IHP assistance and flood insurance enrollment. They have developed a panel dataset of the ZIP codes by utilizing historical data between 2000 and 2011. However, their analysis only considered the availability of federal assistance as a binary variable (IHP assistance was either available or not available). But that is only one part of the picture. Not all ZIP codes receive an equal amount of IHP assistance. In fact, the level of IHP assistance is different in different ZIP codes, which makes the variable continuous. The analysis of the effect of continuous treatment is missing in Kousky et al. (2018b).

Davlasheridze and Miao (2019) have conducted the analysis at the county level using the data between 1998 and 2010. Although their analysis is focused on analyzing the impact of Public Assistance (PA), which compensates the State, Local, Tribal, and Territorial (SLTT) governments to repair, restore, reconstruct, or replace their disaster damaged infrastructure, on the flood insurance enrollment, they have found that Individual Assistance (IA) increased the flood insurance enrollment. It is worth mentioning that PA does not compensate households thus is not directly relevant to flood insurance. Additionally, they have considered all the NFIP participating counties regardless of the Presidential declaration of major disaster. Contrarily, the research presented in this chapter of this dissertation has only considered the NFIP participating counties where a flood related major disaster was declared in a year for a more apples to apples comparison. Thus, the dataset only contained NFIP participating counties that were eligible to receive IHP assistance. Focusing only on these counties, where a Presidential major disaster was declared, also neutralized the effect of political alignment of a county on the likelihood of receiving federal assistance, which has been pointed out by Garrett and Sobel (2003) and Sylves and Búzás (2007). Additionally, this research has used more recent data (between 2016 and 2020) than previous research on this topic.

Table 3.1 also shows the methodology used in existing studies. It can be noticed that the previous charity hazard studies in the U.S. at the macro level have been based on the fixed effects (FE) model. Despite all the merits of this model, it only measures the Average Treatment Effect on the Treated (ATT) (Collischon and Eberl 2020), while the effect of heterogeneity for individual homeowners is not analyzed. In addition to ATT, this research requires the estimation of the Average Treatment Effect (ATE), i.e., the average difference between the treated and the non-treated groups, in this case, the flood insurance enrollment between counties that received the IHP payout and those that did not receive IHP payout despite the declaration of major disaster. Collischon and Eberl (2020) also illustrated that in most cases FE models do not identify true causal effects. Therefore, this research has conducted a macro level analysis using Propensity

Score Matching and Generalized Propensity Score method. Propensity Score Matching is a simple statistical tool that can make more accurate causal inference by balancing non-equivalent groups that might originate from a non-randomized design (Rosenbaum and Rubin 1983). Generalized Propensity Score, which is a modification of the Propensity Score Matching method, is used to analyze the causal effect of a continuous treatment (Hirano and Imbens 2004). Deriving causal inference from observational data with Propensity Score-based method is common (Gianicolo et al. 2020). With propensity score, it is easier to achieve balance of the confounding variables between the control and treatment groups due to the balancing nature of propensity score itself (Hirano and Imbens 2004). Hence, propensity score-based models have been chosen for this research.

3.3 Research Data

This research tests the charity hazard hypothesis, i.e., the availability of post-disaster federal payout in the U.S. reduces the intrinsic motivation of households to insure themselves against flood losses. To test this hypothesis, we have collected all insurance transaction data for five years between 2016 and 2020. The impact of federal payout in a county in the year t is measured against the changes in the flood insurance enrollment in that county in the next year, i.e., year t+1 where $t \in [2016, 2019]$. All collected data was aggregated at the county level for each year to create an unbalanced panel dataset, where the panel unit is county. There were 1294 NFIP participating counties in the dataset where at least one flood related major disaster was declared between 2016 and 2019. After removing counties without any flood damage, we had 1158 datapoints left with positive flood damages where a flood related major disaster has been declared between 2016 and 2019 thus making them eligible to receive IHP. Despite their eligibility, some counties did not receive IHP assistance. The level of assistance was also different among the counties that received IHP assistance from the federal government. There were 445 unique counties in 2016, 374 in 2017, 230 in 2018, and 244 in 2019. It should be noted that this research has only considered counties that are located among the 50 U.S. states and the District of Columbia. Figure 3.1 shows the spatial distribution of the counties considered in this research. It can be noticed that majority of counties are in Texas (156), Missouri (139), Florida (131), North Carolina (120), and Georgia (95).

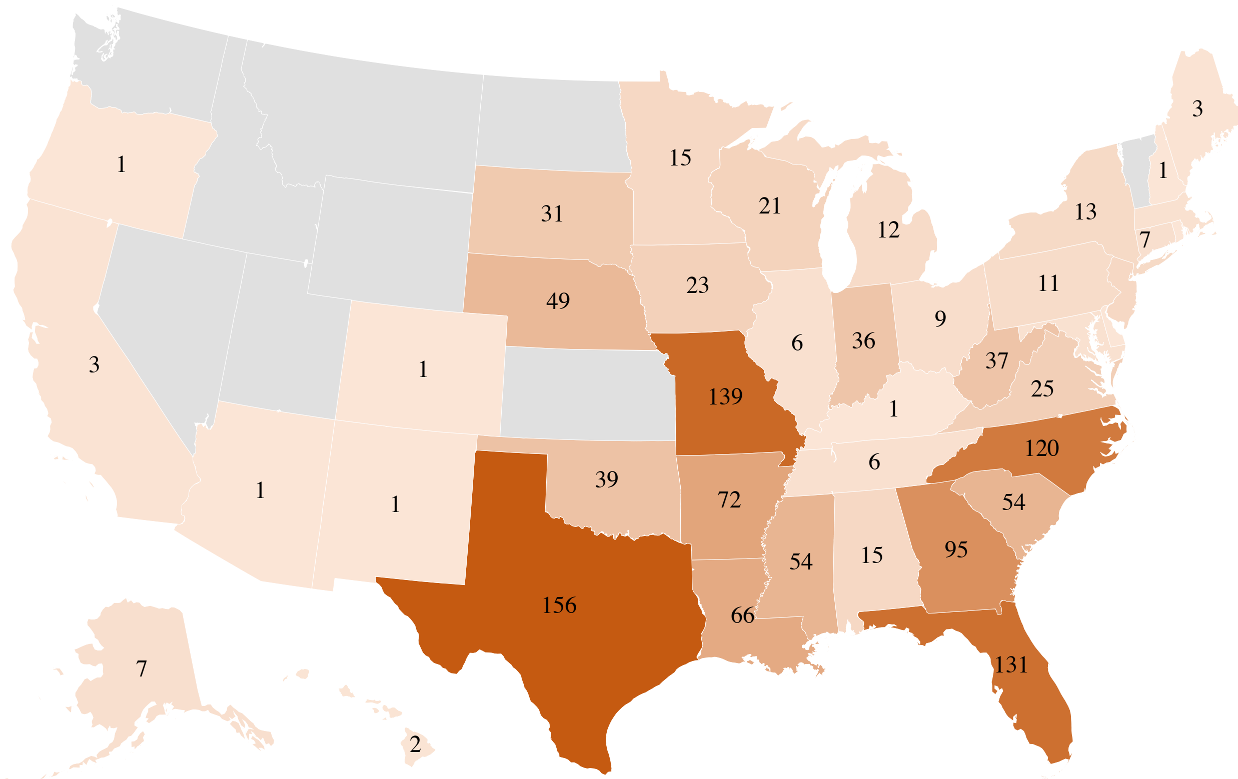


Figure 3.1 Spatial Distribution of the Counties Used in the Analysis

3.3.1 Treatment Variable

Treatment variables are the variables whose impacts are measured, i.e., the federal payouts. In this research, the federal payouts are quantified in terms of two variables (1) IHP Count, i.e., the number of households in a county that received IHP payout in a year t and (2) IHP Amount, i.e., the total amount of IHP support received in that county by those households in a year t where $t \in [2016, 2019]$. Among the 1158 datapoints, 194 datapoints had a zero IHP count and therefore zero IHP amount. This means that those counties did not receive any IHP support from the federal government despite the declaration of major disaster with damages. The other 964 counties received IHP support from the federal government. The level of support is different in different counties, which implies that the treatment variable cannot be considered as binary only. However, when the impact of the IHP support on the flood insurance enrollment is compared between the counties that received IHP support to those that did not, the treatment variables must be considered as binary. On the other hand, to compare the effect of various levels of IHP support among the 964 counties, the treatment variables need to be considered as continuous. Table 3.2 displays the descriptive statistics of the two treatment variables including those with zero values. It is important to note that both treatment variables are highly right skewed with long tails. The number of households that received IHP payout in a county varied between 0 and 182 thousand. On the other hand, the amount of IHP payout in a county varied between \$0 and \$386 million.

3.3.2 Outcome Variables

Outcome variables are the variables on which the effect of the treatment, i.e., the availability of post-disaster federal support in terms of IHP, is measured. There are two outcome variables in this research. They are (1) Percentage NFIP, i.e., percentage change in the number of NFIP policies in a county from year t to t+1 and (2) Percentage TIV, i.e., percentage change in the total insured value (TIV) of those flood insurance policies in that county from year t to t+1, where $t \in [2016, 2019]$. The outcome variables are calculated as in equations 3.1 and 3.2.

$$Percentage\ change\ in\ NFIP\ policies = \frac{(No.of\ policies\ in\ year\ t+1) - (No.of\ policies\ in\ year\ t)}{No.of\ policies\ in\ year\ t} \tag{3.1}$$

$$Percentage\ change\ in\ TIV = \frac{(TIV\ in\ year\ t+1) - (TIV\ in\ year\ t)}{TIV\ in\ year\ t} \tag{3.2}$$

Where $t \in [2016, 2019]$. The impact of the federal payouts is measured by observing the annual changes in the outcome variables in the subsequent year. ıt can be noticed in Table 3.2, which also shows the descriptive statistics of the two outcome variables, that the percentage change is not always positive. In fact, the 25th percentile of the percentage change in NFIP policy count and TIV are both negative. The median percentage change in the NFIP policy count is zero. The average of the percentage change in the number of NFIP policies in the dataset is 5% whereas the average of the percentage change in TIV is 15%. Both outcome variables are right skewed with long tails.

Table 3.2 List of All Variables along with Descriptive Statistics (t ∈ [2016, 2019])

Variable Name	Code	Unit	Variable Type	Time Scale	Mean	SD	Min	25%	Median	75%	Max	Skewness
No. of IHP payout	IHP Count	Count	Treatment	t	956	6939	0	1	22	157	181858	18.1
Amount of IHP payout	IHP Amount	$	Treatment	t	$2,621,293	$16,332,800	$0	$3,053	$99,865	$605,859	$385,796,600	16.0
Percentage change in the number of NFIP policies	Percentage NFIP	%	Outcome	t+1	5%	24%	-100%	-4%	0%	7%	247%	4.0
Percentage change in the Total Insured Value	Percentage TIV	%	Outcome	t+1	15%	58%	-100%	-3%	5%	22%	1072%	11.6
Population	Pop	Count	Covariate	t+1	185914	430843	689	17881	44338	147810	5223719	6.0
Median Household Income	MedIncome	$	Covariate	t+1	50339	13476	20795	40988	48909	57771	121133	1.1
Labor force participation	LabForce	%	Covariate	t+1	58%	8%	24%	52%	58%	64%	75%	-0.5
Occupied housing units	OccUnits	%	Covariate	t+1	83%	9%	39%	79%	84%	89%	97%	-1.4

Table 3.2 continued

Median building value	BuildVal	$	Covariate	t+1	$143,142	$77,786	$0	$94,150	$124,350	$165,975	$944,600	3.3
Population with a bachelor's degree	Bachelors	%	Covariate	t+1	14%	6%	1%	10%	13%	18%	33%	0.7
Flood damage	Damage	$	Covariate	t	$19,761,320	$200,151,900	$12	$62,527	$457,008	$2,694,371	$6,146,881,000	25.8
No. of new mortgages	Mortgage	Count	Covariate	t+1	1080	3028	0	39	146	780	53593	8.2
IHP Eligibility	PerElg	%	Covariate	t	35%	28%	0%	14%	32%	51%	100%	0.7

3.3.3 Confounding Variables

Confounding variables or covariates are the variables that influence both the treatment and outcome variables. If these variables are not controlled properly, they may be unequally present among the comparison groups. As a result, the estimated effect of the treatment on the outcome would be biased. Therefore, it is essential to control these confounding variables. In this research, nine confounding variables have been considered. They are (1) Population of the county in the year t+1, (2) Median household income of the county in the year t+1, (3) Percentage of population in the county that is participating in labor force in the year t+1, (4) Percentage of housing units in the county that are occupied in the year t+1, (5) Median building value of the county in the year t+1, (6) Percentage of population in the county with a bachelor's degree in the year t+1, (7) Extent of flood damage in the preceding year, i.e., year t, (8) Number of new federally backed mortgages in the county in the year t+1, and (9) Percentage eligibility for IHP in the year t.

The extent of flood damage has been calculated as the sum of the flood insurance claim and the flood damage assessed by the FEMA officials for the IHP applicants. The damage assessment considers damage to real property components such as floors, walls, access roads and bridges, electrical, plumbing, HVAC, etc., and personal property components, including appliances, furniture, etc. It should be noted that the assessed damage is higher than the IHP payout in most of the cases as damages that are insured are not eligible for reimbursement through the IHP. Moreover, there is an upper cap for the maximum amount of IHP support that a household can receive. The amount was $35500 for the financial year 2020 (FEMA 2019).

Federal regulations require flood insurance for homes that are located in the 100-year flood zone and have federally backed mortgages. Therefore, the number of new federally backed mortgages in a county in a year can influence the new flood insurance purchase in that year in that county. Hence, the number of new federally backed mortgages in a county has been considered as a confounding variable. The single-family federal mortgage data was collected from the U.S. Federal Housing Finance Agency's website. The data was available separately for two federally backed mortgage lenders – Freddie Mac and Fannie Mae. The number of new federally backed mortgages in a county in a year has been calculated as the sum of the mortgages through Freddie Mac and Fannie Mae.

The final confounding variable is the percentage eligibility or approval rate for the IHP payout in the past. To be eligible to receive IHP assistance, the applicant must demonstrate that (1) the damage is not insured, (2) he or she is a U.S. citizen (or qualified alien), and (3) the damaged property is the primary residence (Kousky and Shabman 2012). Notably, if the applicant lives in a 100-year flood zone and in a community that does not participate in the NFIP, IHP aid is not available to the disaster survivors. This step is taken to increase participation in the NFIP by communities (Kousky and Shabman 2012). Low approval rate among the IHP applicants has been reported by Kousky (2013) and Wu et al. (2017). This lower approval rate can encourage people to purchase flood insurance. On the other hand, it might also discourage potential applicants from applying for federal support. Hence, it has been considered as a covariate. The percentage eligibility has been calculated as the percentage of IHP applicants that were considered eligible to receive IHP support. Table 3.2 lists all the confounding variables along with their time scale.

Next, the correlations between the covariates were tested to avoid multicollinearity. One variable in a highly correlated variable pair, i.e., with pairwise correlation coefficient greater than 0.70, is removed. Figure 3.2 shows the correlation matrix for the covariates. High correlations can be noticed between (1) Mortgage – Pop (0.77), (2) MedIncome – LabForce (0.69), (3) MedIncome – BuildVal (0.75), (4) MedIncome – Bachelors (0.74), (5) LabForce – Bachelors (0.69), and (6) BuildVal – Bachelors (0.72). So, three covariates were removed – Mortgage, MedIncome, and Bachelors. After removing these three covariates, the correlation coefficients among the remaining six covariates were checked again. None of the correlation coefficient was found higher than 0.57. So, the final set of covariates are population, labor force participation, occupied units, median building value, flood damage, and percentage eligibility. Table 3.2 also provides the descriptive statistics of the covariates.

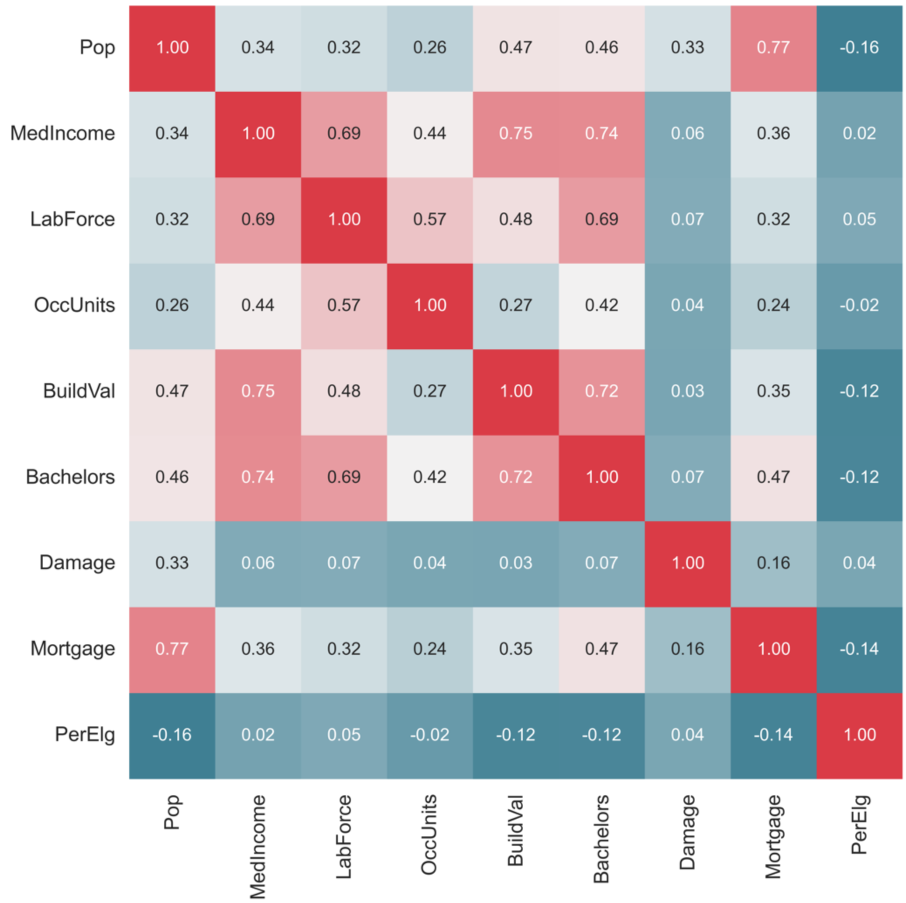


Figure 3.2 Correlation Matrix of the Covariates

3.4 Research Methods

There are two parts to the research. In the first part, the treatment variables are considered as binary. Therefore, the first part compares the differences in the outcome variables between counties that received IHP payout to those that did not receive any IHP payout despite the declaration of flood related major disaster by the U.S. federal government. In the next part, the treatment variables are considered as continuous. The second part develops a dose-response function that demonstrates how different levels of federal support influence the outcome variables. In the second part, the effect of IHP count is measured on the percentage change in the number of NFIP policies and the effect of the IHP amount is measured on the percentage change in the TIV of the NFIP policies.

3.4.1 Part I: Propensity Score Matching for Binary Treatment

Propensity score matching (PSM), which was proposed by Rosenbaum and Rubin (1983), is a quasi-experimental method that uses statistical techniques to create an artificial control group that has similar characteristics of the underlying confounding variables to that of the treatment group. Since the confounding variables influence both the treatment and the outcome variables, controlling them or ensuring their balance, i.e., similar distribution in control and treatment groups is essential to realize the actual effect of the treatment. In this research, the treatment group consists of the counties that received the treatment, i.e., IHP payout whereas the control group consists of the counties that did not receive the IHP payout. The number of counties in the treatment group is 964 whereas the number of counties in the control group is 194. Since the number of samples in the control group is less than that of the treatment group, the 194 counties from the control group are matched against the 964 counties in the treatment group to derive artificial treatment group that has similar characteristics of the covariates as that of the control group.

The framework for analyzing the effect of binary treatment is shown in Figure 3.3. The propensity score is defined as the likelihood of receiving the treatment Z ($Z = 1$ for treatment group and $Z = 0$ for control group) conditional on the observed set of covariates X prior to the application of the treatment. This conditional probability is generally computed using a Logistic Regression model (Logit). It should be noted that the purpose of the Logit model is not to accurately predict whether a county would receive IHP payout or not but to derive propensity score that would ensure the balance of covariates among the treatment group and artificial control group (Austin 2011b). The Logit model computes the likelihood of a county receiving ($Z = 1$) or not receiving ($Z = 0$) federal payout based on the six confounding variables. Before developing the Logit model, the covariates were standardized so that they have zero mean and unit standard deviation. The equation for logistic regression is shown in equation 3.3.

$$Prob(Z = 1) = \frac{e^{\beta_1 X + \beta_2 X^2 + \beta_3 X^3}}{1 + e^{\beta_1 X + \beta_2 X^2 + \beta_3 X^3}} \tag{3.3}$$

Where, Z is the binary treatment variable with two possible values zero and one, X is the vector of covariates, and β_1, β_2, and β_3 are the regression coefficients. It can be noticed that higher order terms (denoted by X^2 and X^3) have been used in the logistic regression. Once, the likelihoods are derived, the logit of the propensity score of the samples in the control group is matched with that of the treatment group to derive the artificial treatment group. Using logit of the propensity score has been recommended by Austin (2011a).

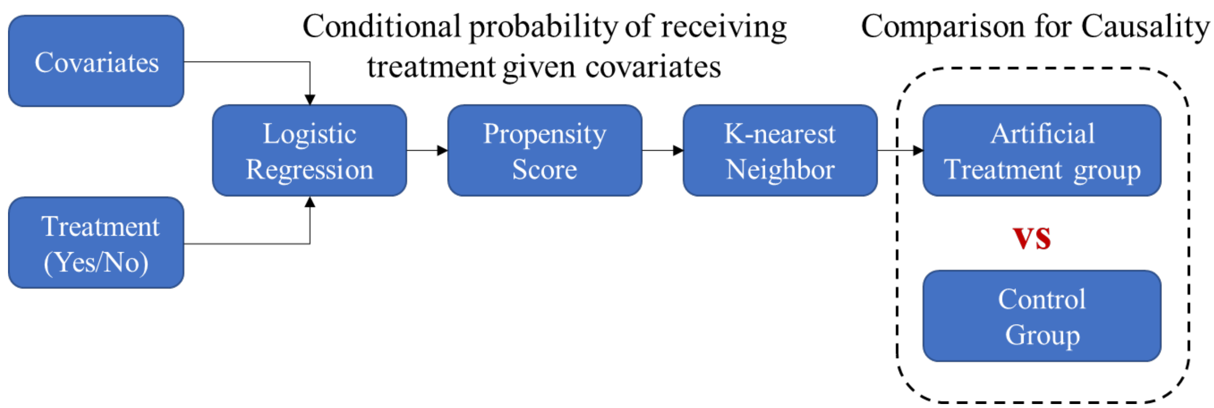


Figure 3.3 Research Framework for Binary Treatment

For creating the artificial treatment group, this research has used one to one matching. In this approach, a sample in the control group is matched to one sample from the treatment group based on the propensity score derived from the Logit model. Among matching techniques, Nearest Neighbor is commonly used (Geldof et al. 2020). In the nearest neighbor algorithm, k = 50 was considered (Geldof et al. 2020). Additionally, as recommended by Austin (2011a), the radius in the nearest neighbor algorithm was computed as 0.2 times of the standard deviation of the logit of the propensity score. After the matching is complete, the artificial treatment group is expected to have similar characteristics of the covariates to that of the control group.

3.4.2 Part II: Generalized Propensity Score for Continuous Treatment

Generalized Propensity Score (GPS) has been defined by Hirano and Imbens (2004) as the conditional density of the continuous treatment given covariates. The definition is shown in equation 3.4.

$$r(z,x) = f_{Z|X}(z|x) \tag{3.4}$$

Where, Z is the treatment variables and X is the vector of measured baseline covariates. The GPS is defined as $R = r(Z,X)$. Imai and van Dyk (2004) refers to the conditional density function $f_{Z|X}$ as the propensity function. The general framework for estimating the dose-response function using the GPS is shown in Figure 3.4. The propensity function is generally estimated by regressing the continuous treatment variable on the set of measured baseline covariates using Ordinary Least Square (OLS) regression. If the continuous treatment variable Z is normally distributed with mean $\beta^T x$ and variance σ^2 (where β and σ^2 are estimated using OLS regression), then the conditional density function can be estimated by the normal density function, i.e., $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\hat{\beta}^T x)^2}{2\hat{\sigma}^2}}$. It is worth mentioning that the calculation of GPS assumes that the continuous treatment follows a normal distribution.

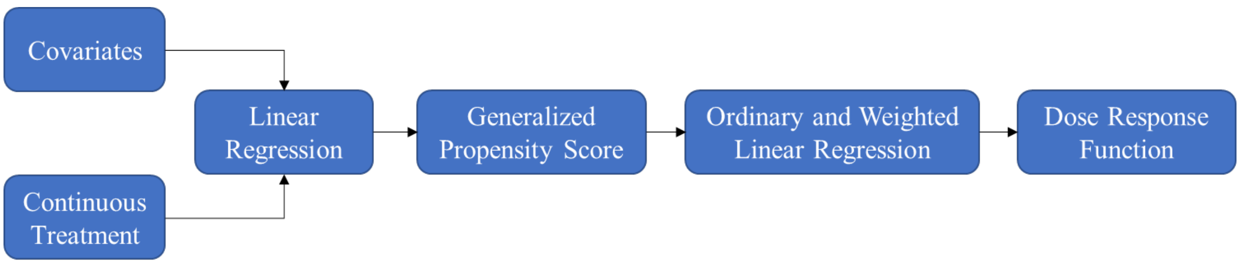


Figure 3.4 Research Framework for Continuous Treatment

The dose response function $\mu(z)$ is defined as equation 3.5 where, $Y_i(z)$ denotes the outcome of the i^{th} sample if it receives a treatment of Z = z. To develop this dose response function, the conditional expectation of Y_i given Z_i and R_i, where R_i is the GPS of the i^{th} sample, is computed. To derive this conditional expectation, the outcome variables should be regressed on the treatment variable and the derived GPS. This research has used two linear regression models: ordinary least square (OLS) and weighted least square (WLS) regression to derive the conditional expectation. Imbens (2000) and Zhang et al. (2016) have suggested that weights could be calculated from GPS that can be further used in estimating the dose-response function. The weight of a sample is calculated as equation 3.6 where $W(Z_i)$ is a function that stabilizes the weights. Zhang et al. (2016) have suggested that the marginal density function of *Z,* as shown in equation 3.7, can be a reasonable choice for *W*.

$$\mu(z) = E[Y_i(z)] \tag{3.5}$$

$$Weight\ of\ sample\ i = \frac{W(Z_i)}{r(Z_i|X_i)} \tag{3.6}$$

$$W(Z_i) = \frac{1}{\sqrt{2\pi\sigma_{sample}^2}} e^{-\frac{(Z_i-\mu_{sample})^2}{2\sigma_{sample}^2}}. \tag{3.7}$$

3.5 Results

As explained earlier, there are two parts of the research. In the first part, the treatment variables were considered as binary. Therefore, the Propensity Score Matching method compared the changes in the outcome variables between the treatment group and artificial control group. In the second part, the Generalized Propensity Score method was used to develop a dose response function that demonstrates the expected changes in the outcome variable for different levels of treatment.

3.5.1 Outcome for Binary Treatment

For the binary treatment, the effect of two treatment variables were estimated together. If no household in a county receives IHP payout, the IHP amount for the county will be zero. On the other hand, if one or more households receive IHP payout, the IHP amount will be more than zero. The reverse is also true. Since the treatment variables are synchronous, they were considered together as one for the binary treatment.

First, a Logit model was developed that expressed the conditional probability of a sample, i.e., a county receiving the treatment given the set of baseline covariates. The logistic regression model used second and third order terms along with the first order of the covariates as shown in equation 3.3. The logistic regression produced a prediction accuracy of 0.82, i.e., for 82% of the counties, the model was able to predict whether one county received the IHP payout or not.

The derived likelihood was utilized to calculate the propensity score, i.e., the logit of the likelihood of receiving IHP payout. This propensity score was further used to derive the artificial treatment set using the Nearest Neighbor algorithm as explained earlier. Initially, the control group had 194 counties and the treatment group had 964 counties. After the matching was conducted, both the control group and artificial treatment group contained 165 counties. Table 3.3 shows the difference of the covariates between the control and treatment group before and after matching.

The differences are expressed in terms of Standardized Mean Difference (SMD) as shown in equation 3.8.

$$d = \frac{\bar{x}_{Treatment} - \bar{x}_{Control}}{\sqrt{\frac{s^2_{Treatment} + s^2_{Control}}{2}}} \tag{3.8}$$

Where d is the SMD, $\bar{x}_{Treatment}$ and $\bar{x}_{Control}$ are the mean of a covariate in the treatment and control group, respectively. $s^2_{Treatment}$ and $s^2_{Control}$ are the sample variance of a covariate in treatment and control group, respectively. SMD is considered small if d is between 0.2 and 0.5, medium if d is between 0.5 and 0.8, and large if d is greater than 0.8 (Cohen 1988).

Table 3.3 Standardized Mean Difference Before and After Matching

Covariate	**SMD Before Matching**	**SMD After Matching**
Population	0.44	0.08
Median Household Income	0.45	0.08
Labor Force Participation	0.32	-0.06
Percentage Occupied Units	0.32	-0.06
Median Building Value	0.54	0.14
Percentage Population with Bachelor's	0.56	-0.06
Flood Damage	-0.14	-0.52
Federally backed Mortgage	0.34	-0.09
Percentage Eligibility	-2.37	-2.60

It can be noticed that for all the covariates except for flood damage and percentage eligibility, the absolute values of the SMD after matching are insignificant, i.e., less than 0.2. For flood damage, the SMD is -0.52, which can be considered small as the absolute value is very close to 0.50. Therefore, the matching process created balance for eight covariates out of nine. The nearest neighbor algorithm matched 165 counties from the control group to 165 counties in the treatment group based on the propensity score to create an artificial treatment group of 165 counties. Next, the outcome variables of the control group and the artificial treatment group were compared to understand the impact of the treatment variables. Figure 3.5 shows the comparison of the outcome variables. From Figure 3.5, the distribution of the two outcome variables looks similar for the control and the artificial treatment groups.

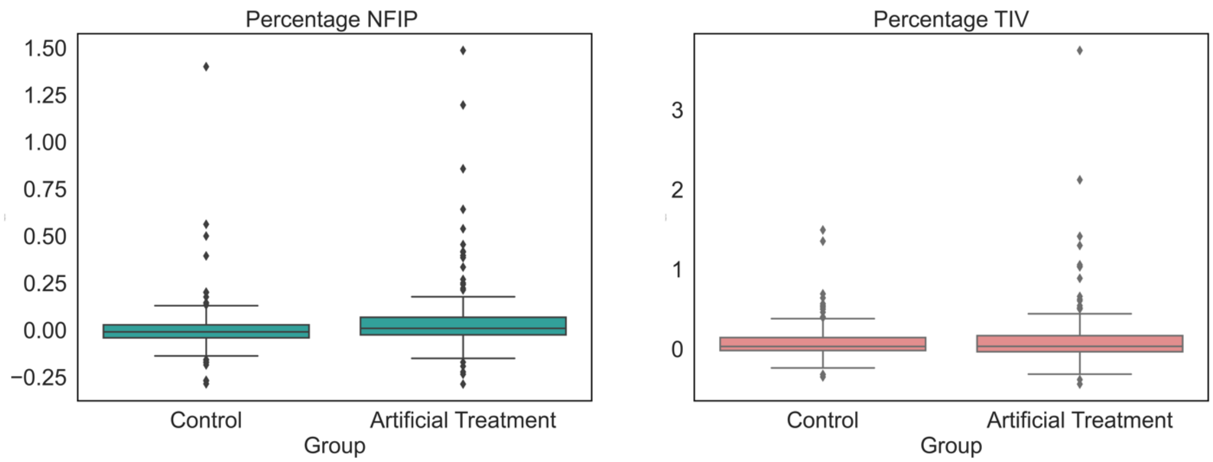


Figure 3.5 Comparison of Outcome Variables

To dig in deeper, the descriptive statistics of the outcome variables were compared. Table 3.4 lists some of the descriptive statistics of the two outcome variables for the control and artificial treatment groups (mentioned as treatment group in Table 3.4). It can be noticed that the average percentage change in the number of NFIP policies in the treatment group is higher than that of the control group. The same is true for the other outcome variable, i.e., the percentage change in the TIV. The standard deviations of the outcome variables are also higher in the treatment group. The results indicate that the counties that have received post-disaster federal support in terms of the IHP payout have purchased more flood insurance in the following year. Therefore, the federal payout did not crowd out the NFIP.

Table 3.4 Intergroup Comparison of Outcome Variables

Variable	**Group**	**Mean**	**SD**	**Min**	**25%**	**50%**	**75%**	**Max**
NFIP	Treatment	5.70%	20.40%	-28.73%	-2.49%	0.94%	6.80%	148.65%
	Control	0.54%	15.04%	-28.57%	-4.07%	-0.95%	2.80%	140.00%
TIV	Treatment	13.30%	42.48%	-44.08%	-3.40%	3.47%	16.88%	374.56%
	Control	8.75%	23.17%	-35.00%	-1.84%	3.21%	14.39%	149.38%

Next, the statistical significance of the intergroup difference was tested. To do that, this research has adopted Kruskal-Wallis test, which is the non-parametric version of the Analysis of Variance (ANOVA) test. The outcome of the test is expressed as *p*-value. If the p-value is less than 0.05, i.e., 5% significance level, the test concludes that the difference is statistically significant. The difference is considered insignificant if the *p*-value is greater than 0.05. The *p*-value for percentage NFIP and percentage TIV were found 0.0008 and 0.99, respectively. The *p*-values indicate that the intergroup difference is significant for the percentage change in NFIP policies whereas it is insignificant for percentage change in TIV. This means that after receiving post-disaster federal assistance, disaster survivors have purchased more flood insurance. Therefore, the difference in the increase of flood insurance policies in counties that received federal assistance was higher than that of the counties that did not receive any federal assistance. However, the increase in percentage of the total insured value is similar between the two groups.

Another way of measuring the impact of the treatment is through the Average Treatment Effect (ATE). ATE represents the average difference in the flood insurance enrollment between counties that received the IHP payout and those that did not receive the IHP payout despite the declaration of major disaster. To derive the ATE, the average of the differences in the percentage change in the NFIP policies and TIV among counties in the control group and their corresponding matched counties in the artificial treatment group are computed. Therefore, it is the average of the difference in the outcome variables of the 165 counties in the control group and artificial treatment group. The ATE for the percentage change in the NFIP policies was found 5.2% and the ATE for the percentage change in the TIV of the NFIP policies was found 4.6%. It can be noticed that the inter-group differences of the means shown in Table 3.4 are the same as the calculated ATEs. This is because the intergroup difference of the means of a variable is mathematically equal to the mean of intergroup differences of that variable. So, the availability of the IHP payout in a county in a year increased the number of NFIP policies by 5.2% with a *p*-value of 0.008 and the TIV of the policies by 4.6% with a *p*-value of 0.99 in the following year.

3.5.2 Outcome of Continuous Treatment

For continuous treatment the effects of the two treatment variables were estimated separately. The effect of the IHP count has been estimated on the percentage change in the number of NFIP policies and the effect of the IHP amount has been estimated on the percentage change in TIV. As shown in Figure 3.4, an OLS regression model was developed first to derive the conditional likelihood of receiving different levels of treatment based on the set of baseline covariates. Six covariates were used for the OLS regression. The regression equation is shown in equation 3.9.

$$z = \beta_0 + \beta_1 X + \varepsilon \tag{3.9}$$

Where z is the continuous treatment variables, β_0 and β_1 are the regression coefficients, and X is the vector of the baseline covariates. It should be noted that OLS regression models were only developed for the 964 counties that had non-zero IHP count and IHP amount, i.e., positive treatment. As explained earlier, the objective of the OLS regression is not to predict the levels of treatment but to derive the generalized propensity score that ensures balance of covariates.

To test the balance of covariates for continuous treatment variable, this research has adopted the method used by Imai and van Dyk (2004). In this method, each covariate is regressed on the treatment variable using OLS regression. Figure 3.6 shows the standardized normal quantile plot of the t-statistics for the coefficients of the IHP Count in each regression. The horizontal lines reflect the critical t-values for a 5% significance level. In the first figure, it can be noticed that 3 points, i.e., 3 covariates are between the horizontal lines. The other six are outside the horizontal lines. This meant that six covariates were highly correlated with the first treatment variable, i.e., IHP count. The other figure is derived identically except that the derived GPS was controlled in each regression. In the second figure, 8 points, i.e., 8 covariates were within the horizontal lines. Only one covariate (mortgages) was outside the two horizontal lines. Therefore, after controlling for the GPS, balancing was achieved for all the covariates except for the mortgages for the first treatment variable, i.e., IHP count.

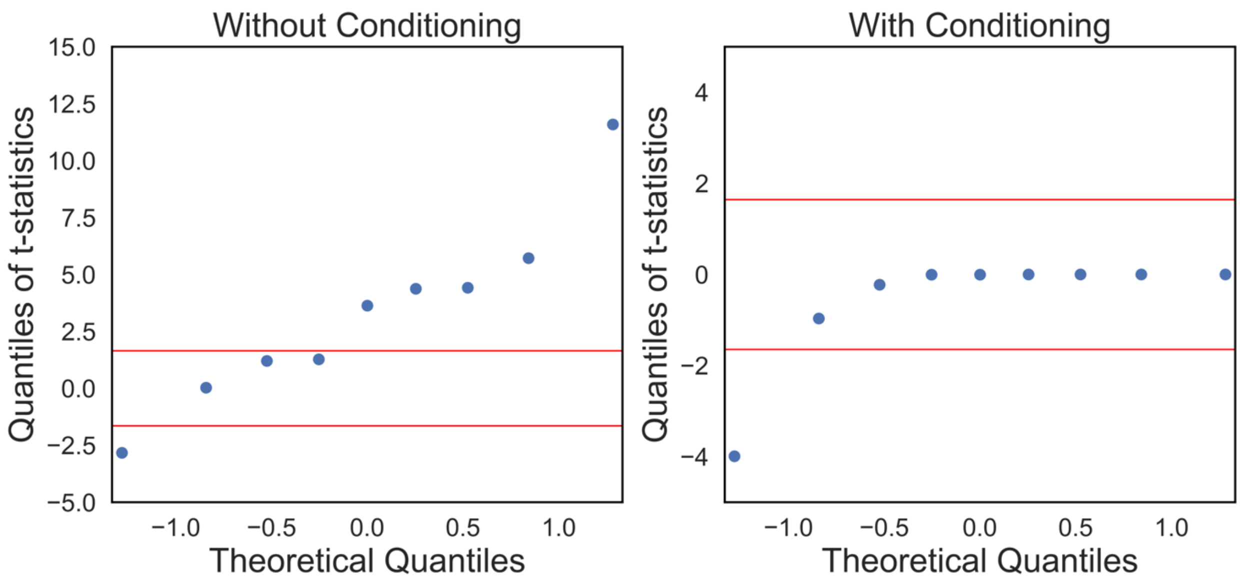


Figure 3.6 Test for Balancing of Covariates for the IHP Count

A similar process was followed to test the balance of the covariates for the other treatment variable, i.e., IHP amount. Figure 3.7 shows the results. The difference in the with and without conditioning is apparent. Without conditioning, only one covariate was within the horizontal lines while two others were on the line. After the conditioning, all nine covariates fell within the two horizontal lines, which indicated the balance of covariates given $\hat{z}$.

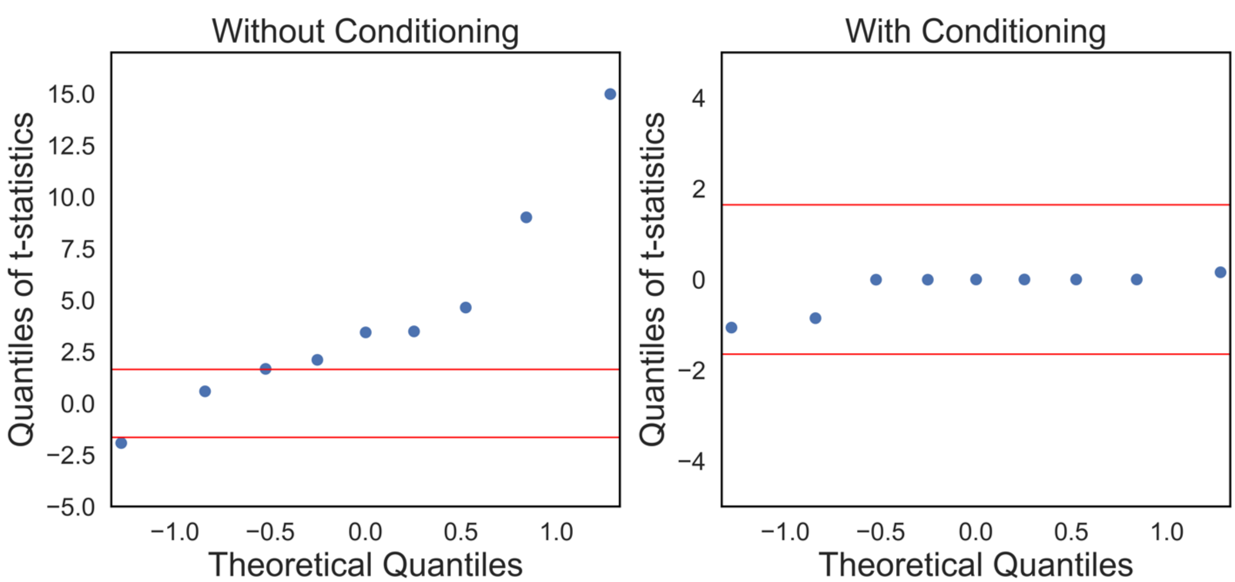


Figure 3.7 Test for Balancing of Covariates for the IHP Amount

For developing the dose response function, the outcome variables were regressed on the corresponding treatment variable and the derived GPS from the linear regression model shown in equation 3.9. The conditional expectations were calculated using two approaches (1) WLS regression using the weights as shown in equation 3.6 and (2) OLS regression without using the weights. The second approach gives equal weights to all the counties whereas the first approach gives more weightage to counties with lesser propensity of receiving treatment given the baseline covariates. Both regressions use an interaction term of the treatment variable and GPS in developing the model (Hirano and Imbens 2004). Table 3.5 shows the regression coefficients for both OLS and WLS models. It can be noticed that regression coefficients have higher significance in WLS models. The models calculated the expected value of the outcome variables for a treatment level $Z = z$ for all the counties. It is important to understand that the coefficients of the treatment variables, i.e., IHP Count and IHP Amount in Table 3.5 are not equivalent to the slope of the dose-response function. This is due to the presence of the interaction terms in the regression equations. The dose response function $\mu(z)$ is the average of all the responses for all the counties. It represents the average impact of the treatment variables on the outcome variables for the samples that received the treatment.

Table 3.5 Regression Coefficients of OLS and WLS for Dose Response Function

	Percentage NFIP				**Percentage TIV**			
	OLS		**WLS**		**OLS**		**WLS**	
Variable	**Coeff**	***p*-value**	**Coeff**	***p*-value**	**Coeff**	***p*-value**	**Coeff**	***p*-value**
Intercept	-0.0041	0.95	-0.0537	0.00	0.16	0.000	-0.02	0.32
IHP Count	6.54×10^{-7}	0.62	2.42×10^{-5}	0.00	N/A	N/A	N/A	N/A
IHP Count×GPS (Interaction Term)	0.68	0.00	0.13	0.44	N/A	N/A	N/A	N/A
GPS	824.49	0.45	1750.12	0.00	1.62×10^{-9}	0.49	4.83×10^{-8}	0.00
IHP Amount	N/A	N/A	N/A	N/A	4.64×10^{-10}	0.69	-3.47×10^{-9}	0.00
IHP Amount×GPS (Interaction Term)	N/A	N/A	N/A	N/A	0.29	0.29	1.46	0.00

Figures 3.8 and 3.9 show the dose response function for the two outcome variables. It can be noticed that the dose response functions are monotonically increasing with the treatment variables. This indicates that higher payout leads to higher flood insurance enrollment. It is true for both outcome variables. It can be seen that the slopes of the dose response function that were derived from the WLS is lower than that of OLS. It has been explained before that the weights used in WLS are inversely proportional to the GPS. Therefore, the WLS method gives less weights to the counties that have higher likelihood of receiving IHP assistance due to the underlying covariates. On the other hand, the OLS gives equal weightage to all the counties. However, both approaches resulted in the same trend of the dose response function. Therefore, the outcomes of research conclude that that the availability of post-disaster federal payout does not crowd out the flood insurance enrollment in the U.S.

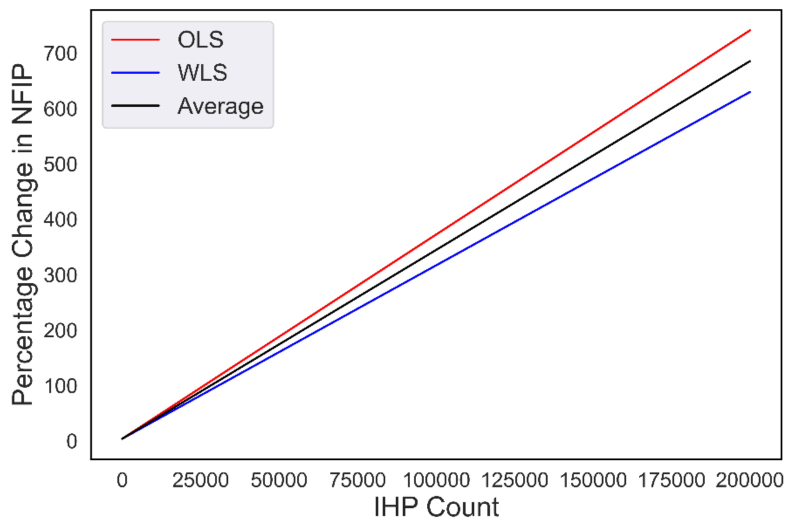


Figure 3.8 Dose Response Function for IHP Count

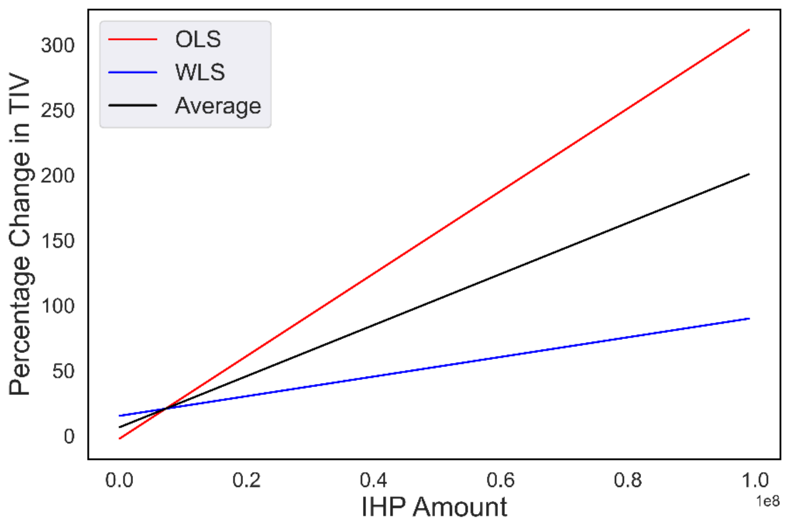


Figure 3.9 Dose Response Function for IHP Amount

To understand the average impact, the dose-response function with OLS and WLS were averaged. The average dose response function is shown in the black line in Figure 3.8 and Figure 3.9. After averaging, it was found that for each 1000 households in a county that received IHP payout, the percentage increase in the number of NFIP policies is 3.41%. On the other hand, for each million-dollar IHP payout in a county, the total insured value of the NFIP policies increased by 1.96%.

3.6 Discussion

Based on the data collected and analyzed, this research has concluded that the availability of the IHP payout increases flood insurance enrollment, i.e., charity hazard does not exist in the U.S. flood insurance market. There are two possible explanations for this research result. First, if a homeowner receives IHP payout to recover from flood losses, he or she is expected to maintain flood insurance to make him or her eligible for future payout. This is a federal requirement (Webster 2019). For homeowners who live in 100-year flood zone and in a community that participates in NFIP, having flood insurance is a requirement for receiving the IHP aid. In absence, FEMA may purchase the flood insurance through the Other Needs Assistance (ONA) funds for 3 years. At its expiration, the applicant must purchase and maintain flood insurance to be eligible for future assistance (Webster 2019). Therefore, this requirement might force the homeowners to purchase flood insurance after receiving the IHP payout.

Second, the insufficiency of the IHP payout may influence the decision of the disaster survivors to insure themselves against future flood losses. The IHP is not designed to compensate all the losses suffered by the affected population. Its purpose is to help the disaster survivors get back on their feet. Therefore, there is an upper cap set on the maximum amount IHP support, which was $35500 for the financial year 2020 (FEMA 2019). However, the insufficiency of the IHP payout has been highlighted multiple times in the past. For instance, Kousky (2013) has investigated the disbursement of IHP after major floods, storms, and tornados in Missouri in 2008. She found that the majority of the aid grants were too small, on the order of a few thousand dollars. Also, more than 50% of the applicants were not granted aid as they were either ineligible or the damage was considered insufficient. The inadequacy of the IHP aid has also been reflected in Sterett (2015).

Figure 3.10 shows the boxplots of the percentage eligibility and the average IHP amount. It can be seen that the median percentage approval rate among IHP applicants in a county is around 37%. The average is 42%. So, on average only 42% of the applicants in a county where a major flood related disaster had been declared were deemed eligible by FEMA to receive the IHP grant. On the other hand, the median of the average IHP amount per household is only $3584 whereas the mean is $4468. This low approval rate along with low average IHP support per household might have influenced the decision of the homeowners and businesses to insure themselves against future flood events, which could have increased the flood insurance enrollment in those counties in the following year.

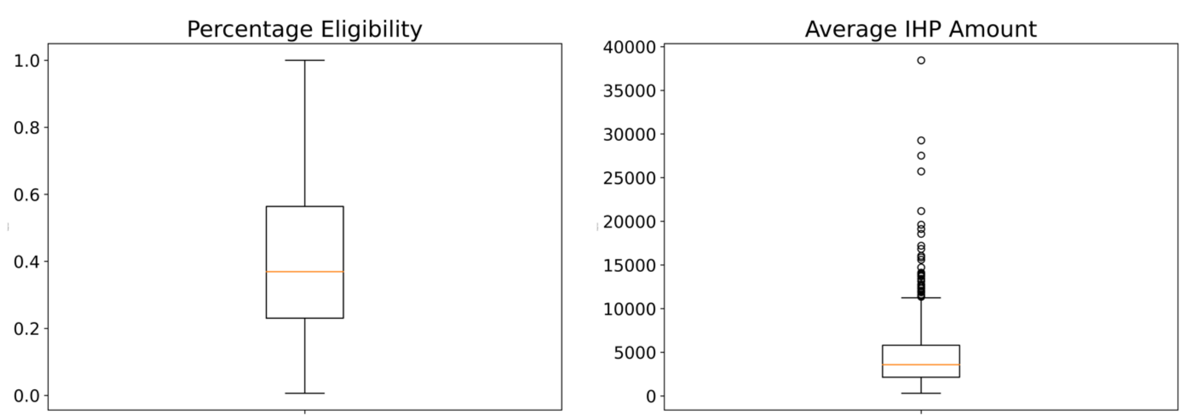


Figure 3.10 Boxplots of Percentage Eligibility and Average IHP Amount

3.7 Conclusion

This research has tested the hypothesis that the availability of the IHP payout crowds out the national flood insurance program in the U.S. This issue is termed as charity hazard. To test the hypothesis this research has collected data, created an unbalanced panel dataset, and used it in two propensity score-based methods. First, the treatment variable was considered as binary to compare the flood insurance enrollment in counties that received IHP payout to those that did not receive any IHP payout despite the declaration of major disaster. It was found that the availability of the IHP payout in a county in a year increased the number of NFIP policies by 5.2% and the TIV of the policies by 4.6% in the following year. Next, the treatment variable was considered as continuous to estimate the impact of receiving different levels of IHP payout on the outcome variables. It was found that for each 1000 households in a county that received IHP payout, the percentage increased in the number of NFIP policies was 3.41%. On the other hand, for each million-dollar IHP payout in a county, the total insured value of the NFIP policies was increased by 1.96%. Therefore, this research has concluded availability of post-disaster federal payout increases the flood insurance enrollment in the U.S, i.e., charity hazard does not exist in the U.S. flood insurance market. The existing federal regulations for IHP and the inadequacy of the IHP payout are possible causes for this.

4. FLOOD RISK FACTORS AND FLOOD INSURANCE PAYOUT

Abstract

This chapter presents a regression model that quantifies the causal relationship between flood risk factors and the flood insurance payout in the U.S. The flood risk factors that have been considered in this research are flood exposure, infrastructure vulnerability, social vulnerability, and the number of mobile homes. Historical data for the annual flood insurance payout, flood risk factors, and other control variables were collected for six years between 2016 and 2021 and used in a Mixed Effects Regression model to derive the empirical relationships. The regression model expressed the natural logarithm of the annual flood insurance payout in a county based on the flood risk factors and control variables. The paper presents the regression coefficients that quantify the causal influence. It has been found that all four flood risk factors have statistically significant positive influence on the flood insurance payout in a county. However, the extent of the influence is different for different flood risk factors. Among them, flood exposure has the highest influence on the flood insurance payout, which is followed by the number of mobile homes, infrastructure vulnerability, and social vulnerability. Since the federal flood insurance program in the U.S. is under huge debt to the U.S. treasury, the government should plan for effective risk reduction that can reduce the flood insurance payout in future to keep the program solvent. The outcomes of this research are expected to facilitate that decision-making process by providing the empirical relationship between flood risk factors and flood insurance payout.

4.1 Introduction

The National Flood Insurance Program (NFIP), which is run by the Federal Emergency Management Agency (FEMA), started in 1968 under the National Flood Insurance Act. The reluctance of private insurers to provide flood insurance created the need for the NFIP (Kousky et al. 2020). Purchasing flood insurance was voluntary till 1973. After 1973, buying flood insurance was mandated for properties with mortgage from a federally regulated or backed lender that are located in a NFIP participating community and within 100-year flood zone by the Flood Disaster Protection Act of 1973. While the Federal government offers flood insurance to the households, it has been observed in the aftermath of the flood related disasters that most of the sufferers are uninsured or underinsured (Kousky 2011). A congressional research report published in 2019 (Horn 2019) showed the flood insurance take up rate for some of the recent flood events. The report recorded the average NFIP take up rate in the 100-year flood zones for multiple flood events such as the South Carolina Flood in 2015 (30%), Louisiana Flood in 2016 (31%), Hurricane Harvey in Texas (21%), Hurricane Irma in Florida (31%). Munich Re reported in 2020 that there were 14.6 million properties in the U.S. that were at substantial flood risk, i.e., located in 100-year flood zone. However, historical records showed that in 2020, there were approximately 4 million active NFIP policies in the U.S., which also demonstrates the low take-up rate of flood insurance in the U.S. Moreover, it has also been found that NFIP policies are often short lived, i.e., they do not get renewed (Michel-Kerjan et al. 2012).

NFIP had a cumulative debt of $20.5 billion to the U.S. Treasury as of 2020 after the federal government forgave $16 billion debt in 2016 (Grigg 2019, Horn 2020). The debt could be partly attributed to the low take-up rate. Additionally, it was projected that the annual deficit of collected premium and the expected payout would remain $1.4 billion in future (CBO 2017). This problem cannot be solved simply by raising the cost of insurance premiums. Previous researchers have found that the price elasticity of the demand for flood insurance is inelastic, which means that the demand for flood insurance is relatively insensitive to the price (Browne and Hoyt 2000, Landry and Jahan-Parvar 2011). However, with a higher cost of flood insurance premium, NFIP might end up with the adverse selection problem, where only the households with high flood exposure purchase the flood insurance. The adverse selection problem arises due to the presence of asymmetrically used information between insurer and insured. Bradt et al. (2021) have found substantial evidence that confirms the occurrence of asymmetrically used information for flood insurance under NFIP. The adverse selection problem is expected to increase the probability of future payouts. On the other hand, NFIP as a government sponsored program is the insurer of last resort even for the households that are deemed uninsurable by private flood insurers (FEMA 2015, Horn and Webel 2021). Additionally, one of the long-term goals of NFIP is to plan for effective flood risk reduction to minimize flood insurance payouts (Horn and Webel 2021). For that, understanding the empirical relationship between flood related risk factors, i.e., the factors that influence the flood risk and subsequently the flood insurance payout is essential. Moreover, climate change is expected to increase the frequency and intensity of natural hazards. This poses another challenge on the long-term financial viability of the NFIP without any risk reduction and resilience initiative. All these issues regarding the NFIP create a need for an improved understanding of the factors that influence the flood insurance payout. Identifying the relationships between those factors and the flood insurance payout and quantifying them can help the policy makers designing effective policies such as property buyout, strengthening infrastructure, etc., that can reduce the future flood insurance payout, which is the objective of this research.

To do that, this research has developed an empirical model that quantifies the causal relationship between flood insurance payout and different flood related risk factors (henceforth referred as flood risk factors) such as flood exposure, infrastructure vulnerability, social vulnerability, community resilience, number of mobile homes, etc., by using historical data for six years between 2016 and 2021 in a Linear Mixed Effects Regression model. Although it is known that the identified flood risk factors influence the flood insurance payout, the proposed model quantifies that causal relationship while considering necessary control variables. The outcomes of the model demonstrate positive causal relationships between flood insurance payout and four flood risk factors, i.e., flood exposure, infrastructure vulnerability, social vulnerability, and number of mobile homes.

4.2 Research Background

The NFIP has been considered problematic almost since its inception. Schilling et al. (1987) found that the program had been largely unsuccessful in the coastal areas due to paying more on claims than the collected premiums. NFIP has been using flood maps for determining the flood insurance premiums for households and businesses. The U.S. Department of Homeland Security (DHS) reported in 2017 that only 42% of those flood maps can adequately identify flood risk, i.e., majority of them were outdated and could not reflect the true flood risk of a property (DHS 2017). Payment of flood insurance claims is one of the key operating expenses of the NFIP. Historical records showed that between 1978 and 2017, NFIP collected \$60 billion in premiums while paid \$65 billion as payouts (Grigg 2019). In 2016, the U.S. congress had to forgive \$16 billion debt of NFIP to maintain its solvency (Grigg 2019). The program also pays one-third of its income from the collected premiums to the financial intermediaries for underwriting the flood insurance policies although none of the flood risk is borne by these intermediaries (Grigg 2019). Repetitive loss properties account for 25-30% of the claims although they are only about 1% of the insured properties (Grigg 2019).

Researchers have long been recommending reforms in NFIP. Michel-Kerjan and Kunreuther (2011), Akabas (2014), McShane and Wie (2019) have recommended risk-based premiums, unsubsidized rates, protection of low-income groups, forgiveness of the debt to the U.S. Treasury, reduction of exposure by reinsurance and CAT bonds, etc., for reforming the NFIP. However, affordability of the NFIP policies remains one of the key challenges for the lower-income households (Shively 2017). Frazier et al. (2020) claimed that an NFIP reform without any considerations for the socioeconomic vulnerability will create barriers for lower-income residents. On the other hand, Wagner (2022) has discovered that the willingness to pay for flood insurance is remarkably low in the U.S.

As explained earlier, in the past FEMA used flood maps, which did not always reflect the true flood risk of a property (DHS 2017). In 2021, FEMA brought a new approach named Risk Rating 2.0 that calculates the flood insurance premium of a household or business based on its true flood exposure thus making it more accurate. All new policies from October 1, 2021, onwards are subject to this new risk rating method. FEMA sees this new approach as a transformational leap forward that will set flood insurance premium rates fairer and more equitable. FEMA estimated that Risk Rating 2.0 will immediately decrease the monthly flood insurance premium for 23% of the policy holders. While 66% of them will see their monthly flood insurance premium increase by less than $10. The remaining 11% will face a monthly increase of flood insurance premium by more than $10 (FEMA 2022). However, there has been some controversy regarding this Risk Rating 2.0, as it is predicted that some states in the U.S. such as Louisiana could see an increase in premium for more than 80% of the existing policies (Murphy 2022).

It has been explained earlier that NFIP's long term solvency and financial issues cannot be simply solved by raising the flood insurance premiums as it might further reduce the demand for flood insurance. The reduced demand could increase the extent of uninsured losses from floods in future. On the other hand, it is expected that the frequency and severity of natural hazards would increase in the long term due to climate change (Smith 2023). Therefore, it is essential that NFIP plans for flood risk reduction to keep the program financially viable. Other than providing flood insurance to households, NFIP has a long-term objective to reduce federal expenditures on post-disaster assistance (Horn and Webel 2021). Therefore, NFIP is expected to ensure that future payouts are kept within limits. This requires the understanding of the causes that influence the annual NFIP payout, i.e., a causal model that explains the flood insurance payouts based on different flood risk factors so that appropriate flood risk reduction strategies and/or policies can be planned to mitigate the impact of those flood risk factors on the NFIP payouts.

In the last two decades a mixed methodology of statistical methods, data analytics, and machine learning techniques have become more prominent in the flood risk reduction studies (Spekkers et al. 2014, Sadler et al. 2018, Desai and Ouarda 2021). Researchers have been using historical data to develop various types of empirical models to derive insights from those data. Similarly utilizing historical NFIP payout data to derive empirical relationships between flood risk factors and NFIP payout can be one of the possible ways to enhance the understanding of NFIP payout. In recent years researchers have been exploring the use of redacted flood insurance claims data. For instance, Wing et al. (2020) have used historical NFIP redacted claims data to derive several insights on flood depth-damage functions. They have found that the observed flood losses are a non-monotonic function of the flood depth. Mobley et al. (2021) have utilized the NFIP redacted claims data to develop a continuous flood hazard map for the Texas Gulf Coast region using a Random Forest model. Knighton et al. (2020) have also used historical flood insurance claims data to develop a model that can predict the number of parcel-level and census tract-level flood insurance claims in New York state using Random Forest classification and regression model. While their model is useful, it was only developed for the New York state. Ghaedi et al. (2022) have also utilized NFIP redacted claims data to predict the extent of flood losses in terms of the number of flood insurance claims based on different factors such as flood peak ratio, Giovanni flooded fraction, land slope, population per area, number of NFIP policies, etc. While the number of flood insurance claims is an important proxy for flood damage, it does not reflect the complete picture as their research did not consider other flood risk factors such as infrastructure vulnerability, social vulnerability, existing resilience, etc. These factors influence the extent of flood damage (Choi et al. 2019, Sanders et al. 2020, Koc and Işık 2021), which influences the NFIP payout.

Therefore, the reviewed literature suggests that flood risk reduction is essential to keep the NFIP financially stable in the long term especially due to the increasing intensity of flood hazard caused by climate change. For planning effective flood risk reduction, empirical models that can demonstrate causal relationships between flood risk factors and flood insurance payout are required. To the best of the authors' knowledge, that causal model is yet to be developed. The objective of this research is to fill that research gap by developing a data-driven causal model that explains flood insurance payout based on different flood risk factors while considering possible control variables.

4.3 Flood Risk Factors

Risk factors are common in clinical science and defined as the factors that increase the likelihood of developing a disease. Similarly, this research has defined flood risk factors as the factors that increase the likelihood of flood losses and subsequently the flood risk in a region. Since flood risk factors influence the extent of flood loss, it is safe to assume that they influence the extent of flood insurance claims due to flood loss. Additionally, this research has narrowed the focus only on the factors that can be controlled through human interventions, which has led to five controllable factors that influence the extent of flood losses in a region. They are (1) flood exposure, (2) infrastructure vulnerability, (3) social vulnerability, (4) community resilience, and (5) the number of mobile homes.

4.3.1 Flood Exposure

The flood exposure measures the representative value of buildings exposed to river and coastal floods (Zuzak et al. 2021). It is evident that higher flood exposure leads to increased likelihood of higher flood losses. Therefore, it has been considered as a flood risk factor. It is worth mentioning that this research did not calculate the flood exposure. The data was collected from FEMA's National Risk Index, which records the flood exposure for all the counties (Zuzak et al. 2021). The exposure is initially computed at the census block level and then aggregated to the county level. To estimate the exposure to river and coastal flood hazards, the hazard occurrence and susceptible zone polygons (as suggested by FEMA) are overlapped with appropriate administrative areas (county for this research). The resulting intersecting shape measures the area of exposure. The number of buildings within that area is the measure of flooding exposed buildings. The detailed procedure for the exposure calculation can be found in FEMA's National Risk Index's Technical Documentation (Zuzak et al. 2021).

For coastal flood exposure calculations, the susceptible zone polygons included 100-year flood zones, 500-year flood zone, and the National Oceanic and Atmospheric Administration's (NOAA) minor, moderate, and major High Tide Flooding (HTF) zones (Zuzak et al. 2021). HTF occurs when sea level rise combined with other local factors push the water level above the high tide mark. The HTF falls under three categories: minor (when the water level touches 1.8 feet above the high tide), moderate (when the water level reaches 2.8 feet above the high tide), and major (when the water level touches 3.9 feet above the high tide). For river flood exposure calculations, the susceptible zone polygons included the 100-year flood zones (Zuzak et al. 2021). The combined flood exposure was calculated as the sum of river flood exposure and coastal flood exposure. It should be noted that the data on flood exposure is only available for one year. The values are in 2022-dollars. Therefore, the exposure has been assumed constants for all six years between 2016 and 2021 due to lack of better data. Figure 4.1 shows the log transformed flood exposure of different counties. The transformation is done to make the differences apparent. It can be noticed that the counties with the highest flood exposure are primarily located in the coastal regions. These counties are exposed to coastal floods, which have higher susceptible zone polygons, thus increasing the exposure value.

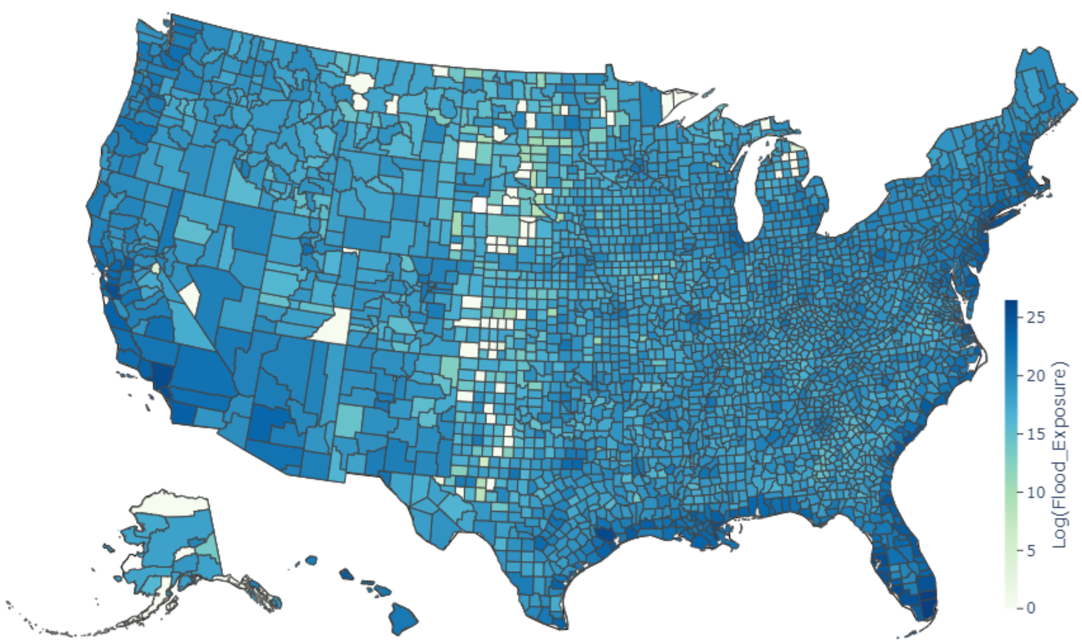


Figure 4.1 Flood Exposure of Counties (Log Transformed)

4.3.2 Infrastructure Vulnerability

Infrastructure works as the first line of protection against natural hazards such as floods, severe storms, hurricanes, etc. As a result, they often fail due to suffering from physical damage during these natural hazards. Ezell (2007) defined vulnerability as the susceptibility to failure under threat scenarios such as natural hazards, intentional attacks, etc. Evidently, vulnerable infrastructure leads to severe flood losses and subsequently higher flood insurance claims from the households. Therefore, it is essential to consider infrastructure vulnerability as a predictor in the causal model.

The U.S. federal government compensates the disaster affected state, local, tribal, and territorial governments to repair, restore, reconstruct, or replace their disaster damaged infrastructure through the FEMA managed Public Assistance (PA) program. Through this program, state and local governments repair disaster damaged roads and bridges, utilities, water control facilities, and other types of infrastructure. Bhattacharyya et al. (2023) have recently used that historical PA data in understanding the vulnerability of roads and bridges in the U.S. They have found that the poorly maintained roads and bridges have failed more often and thus required more federal assistance for repairing. Based on that research outcome, this research has considered the per capita PA amount received by a county in a year as a proxy variable representing the vulnerability of infrastructure in that county. Figure 4.2 shows the log transformed per capita total PA payout for the counties between 2016 and 2021. It can be noticed that the counties on the eastern half of the U.S. have received more PA assistance for flood related disasters. It is because these counties are more exposed to flood hazards than those in the western half. However, counties in California, Oregon, and Washington have received PA funding between 2016 and 2021 as they are exposed to coastal flood hazard.

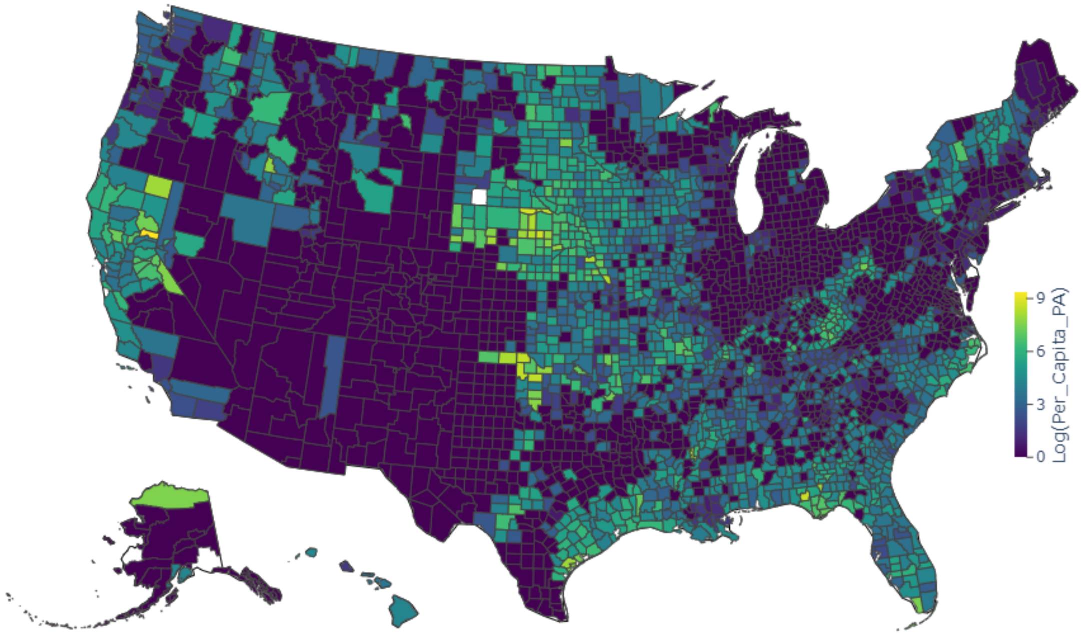


Figure 4.2 Per Capita Public Assistance Payout in Counties (Log Transformed)

4.3.3 Social Vulnerability

Social vulnerability includes the socio-economic and demographic factors that increase or reduce the impacts of natural hazards on a community (Tierney et al. 2001, Heinz Center 2002). Previous researchers have found that the impact of disasters is more prevalent in the socially vulnerable population. For instance, Campbell et al. (2020) have found that vulnerable populations suffer the most damage from floods. These include seniors, people with functional and access needs, people of lower economic status, and other minorities. Several research works have analyzed the role of multiple socio-demographic factors on flood vulnerability and have found significant impact (Cutter et al. 2003, Zhang and You 2014, Dandapat and Panda 2017, Emrich et al. 2020, Drakes et al. 2021, Koc and Işık 2021).

This research utilizes the Social Vulnerability Index (SVI) proposed by the U.S. government's Centers for Disease Control and Prevention (CDC). The index is developed by utilizing 16 socioeconomic variables such as poverty level, unemployment, health insurance, race, ethnicity, disability, elderly population, etc. The SVI data ranges between zero and one, where zero and one are the lowest and highest level of social vulnerability, respectively. The SVI data for each county, which was collected from CDC's website, was only available for the years 2016, 2018, and 2020. Therefore, the SVI for 2017 and 2019 were calculated by interpolating, i.e., by taking average in this case. Additionally, when this research was conducted, the SVI for 2021 was not published. So, the SVI for 2021 was assumed to be equal to that of 2020. Figure 4.3 shows the spatial distribution of the average annual SVI of the counties.

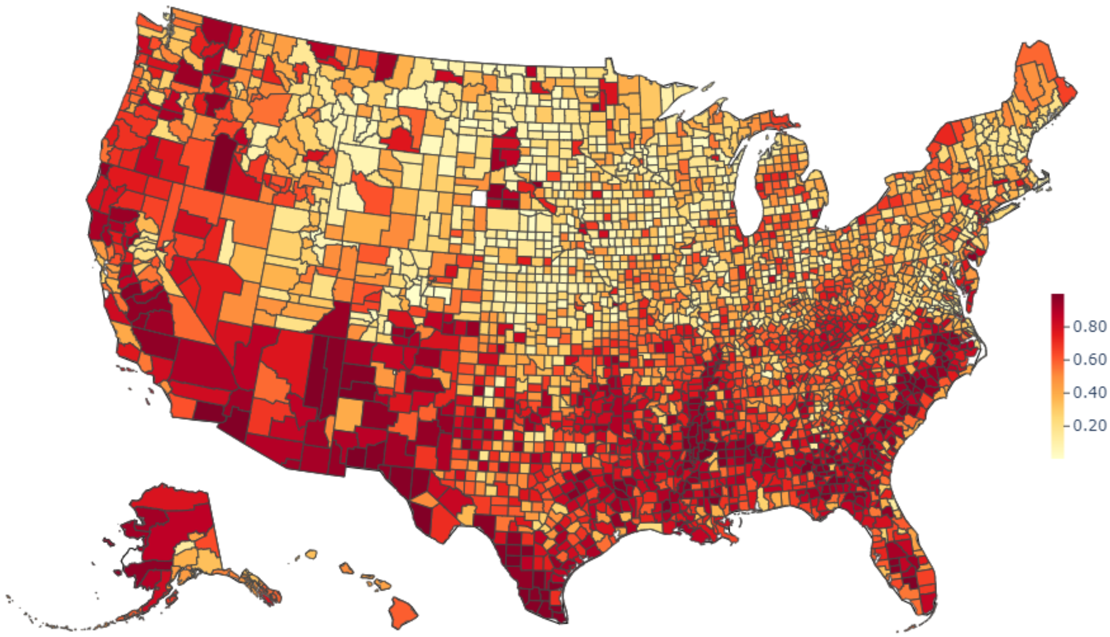


Figure 4.3 Average Annual SVI of Counties

It can be noticed that the southern states are more socially vulnerable than the northern states. As a result of the higher social vulnerability, the southern states are more at risk of flood losses. At the same time, these counties are also exposed to coastal floods, which increases the likelihood of high flood loss compared to the northern states.

4.3.4 Community Resilience

The National Institute of Standards and Technology (NIST) has defined Community Resilience as the ability of a community to prepare for anticipated natural hazards, adjust to changing conditions, and withstand and recover speedily after the disaster (NIST 2020). The existing coping capacity of a community is an important predictor for estimating the impact of a natural hazard in that community (Scheuer et al. 2011, Yang et al. 2013, Terti et al. 2015). Choi et al. (2019) have proposed that a disaster resilient community needs capacities in its all seven layers of critical infrastructures. These seven layers are civil, civic, social, educational, financial, environmental, and cyber. However, this research has utilized the Baseline Resilience Indicator for Communities (BRIC) developed by University of South Carolina's Hazards and Vulnerability Research Institute because of the availability of relevant data (Cutter et al. 2014). The indicator is developed based on 49 factors representing six types of resilience: social, economic, community capital, institutional capital, housing or infrastructure, and environmental (Cutter et al. 2014). The community resilience data was also collected from FEMA's National Risk Index database for the year 2020. The community resilience has been assumed constant for other years due to lack of available data. The community resilience values ranged between zero and one hundred. Zero represents the lowest level of community resilience whereas one hundred represents the highest level of community resilience. Figure 4.4 shows the spatial distribution of the community resilience. It can be noticed that the spatial distribution of social vulnerability and community resilience are inversely correlated. The Midwestern and Northeastern states that have low social vulnerability have higher community resilience. On the other hand, the Southern states that have higher social vulnerability have lower community resilience.

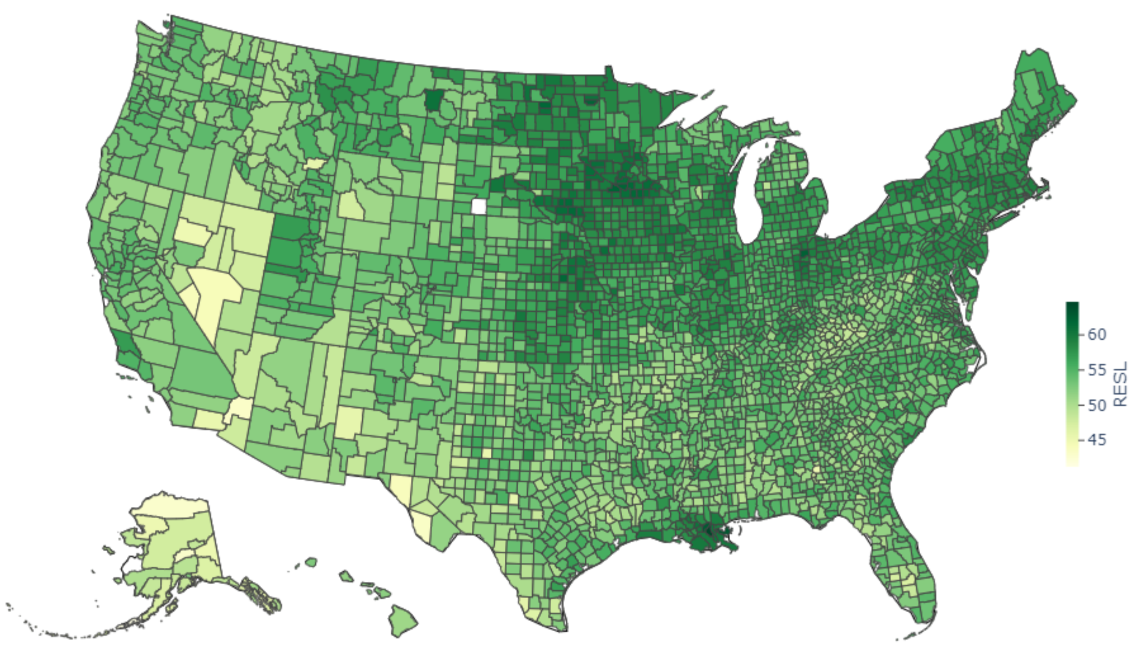


Figure 4.4 Community Resilience of Counties

Despite the popularity, index-based approach for quantifying community resilience has recently received some criticisms. For instance, Yabe et al. (2022) have listed some of the major limitations such as failing to account for complex intercedences between coupled socio-technical systems and dynamic process of resilience, difficulty to validate and test, etc., of using static index for quantifying community resilience. Yabe et al. (2021) have proposed using more dynamic approach like utilizing human mobility data to model the recovery of a disaster affected community. However, this type of approach for modeling community resilience is event specific and does not provide a measure of the existing capacities in a community in absence of a disaster. Moreover, this research required having an annual county level measure of community resilience for 1813 counties, which is not feasible with the approach explained in Yabe et al. (2021) and Yabe et al. (2022). Hence, BRIC has been chosen as a measure of the existing coping capacities in a county to overcome disasters. Future research could be designed to account for county-based approach of resilience that could integrate micro-level dynamic measures of community resilience.

4.3.5 Mobile Homes

The U.S. Department of Housing and Urban Development (HUD) defines mobile homes as structures that are assembled in a factory and shipped in one or more sections on a permanent chassis. These homes may or may not have permanent foundations. As a result, they are vulnerable to wind and water forces despite the recent advancements in making them resilient to strong wind and water forces (Baker et al. 2014, Rumbach et al. 2020). The U.S. Census Bureau's records show that approximately one third of the existing mobile homes are located in Florida, Texas, North Carolina, and California, i.e., the states that are exposed to coastal floods. Moreover, mobile home communities are often dominated by low- and moderate-income households, who are less likely to be prepared to face disasters caused by natural hazards (MacTavish and Salamon 2001, Aman and Yarnal 2010, Baker et al. 2011). Additionally, Shen (2005) analyzed the locations of mobile homes in rural North Carolina and found that they are twice as likely to be located in 100-year flood zones than other forms of housing. Due to all these factors, the number of mobile homes in a county has been considered a flood risk factor in this research.

4.4 Research Data and Methods

As explained earlier, the objective of this research was to develop a causal model that can explain the flood insurance payout in a county based on the identified flood risk factors while considering the necessary control variables. Towards that objective, data from multiple publicly available sources such as FEMA, U.S. Census Bureau, the National Oceanic and Atmospheric Administration (NOAA), the Center for Disease Control and Protection (CDC), etc., were collected for six years between 2016 and 2021. All the collected data were aggregated at the county level to develop an unbalanced panel dataset with county as the panel unit. Table 4.1 lists all the variables used in this research.

4.4.1 Response Variable

The response variable for the causal model is the annual NFIP payout in a county between 2016 and 2021. It should be noted that this research has only considered the NFIP participating counties where the number of active NFIP policies and the annual NFIP claims are at least one. This filtering resulted in an unbalanced panel dataset since all the counties did not have positive flood insurance claims each year. As a result, the final unbalanced panel dataset had 4327 datapoints representing counties from 48 states that excluded Alaska, Hawaii, and Washington D.C. There were 1813 unique counties in the final dataset, where the response variable ranged between \$8 and \$4.43 billion. Additionally, the response variable, i.e., the annual NFIP payout was extremely right skewed with the skewness coefficient of approximately 55, which is very high for linear regression purposes. Therefore, the response variable was transformed to its natural logarithm. After the transformation, the response variable became more symmetric than before and the skewness coefficient was down to 0.56, which is approximately 99% reduction. Therefore, this log transformed annual NFIP payout has been used as the response variable for developing the causal model. Table 4.1 presents other descriptive statistics of the response variable.

Table 4.1 List of Variables with Descriptive Statistics

Variable	Variable Code	Unit	Data	Supporting Literature	Mean	SD	Min	25%	50%	75%	Max
Annual NFIP Payout	NFIP	$	2016-21	-	3,729,738	71,774,040	8	25,015	85,881	357,422	4,428,741,000
Flood Exposure	ExposureB	$	2022	Stephenson and D'ayala (2014), FEMA (2021)	5,774,682,000	21,675,160,000	-	199,038,400	536,000,000	2,513,558,000	313,000,000,000
Infrastructure Vulnerability	Infra	$	2016-21	Len et al. (2018), Sanders et al. (2020)	27	184	-	-	-	3	5,745
Social Vulnerability	SOVI	NA	2016, 2018, and 2020	Cutter et al. 2003; Zhang and You 2014	0.55	0.27	0.00	0.32	0.57	0.79	1.00
Community Resilience	RESL	NA	2020	Cutter et al. (2014), Choi et al. (2019)	55.10	2.57	44.08	53.39	55.19	56.96	64.67
Mobile Homes	MobileHomes	Count	2016-21	Baker et al. 2014, Rumbach et al. 2020	4999	7704	0	1412	2835	5523	90970
Rainfall	Anomaly	Inches	2016-21	Tarhule (2005), Zhang et al. (2018)	8.10	8.29	-18.88	2.08	7.37	13.68	42.73
Flood Damage	Damage	$	2016-21	Botzen et al. (2009), Boamah et al. (2015)	14,291,720	281,958,400	-	-	-	100,000	10,000,510,000

Table 4.1 continued

No. of NFIP Policies	Policy	Count	2016-20	Owusu-Ansah et al. (2019), Moreira et al. (2021)	4442	16896	5	149	465	1969	324898
Total Insured Value	TIV	$	2016-20	Patankar and Patwardhan (2016), Wang and Sebastian (2021)	1,906,414,000	7,932,472,000	811,928	36,214,930	145,480,600	762,326,300	163,000,000,000
Population Density	PopDensity	Count per Sq. Miles	2016-21	Suriya et al. (2012), Santos et al. (2018)	574.6	2311	1.5	47.9	117.8	381.3	71905.1
Median Building Value	MedBValue	$	2016-21	Schröter et al. (2014), Wing et al. (2020)	3,692,616,000	26,855,680,000	39,500	118,750	163,000	263,700	951,000,000,000
Percentage of Occupied Buildings	PerOccupied	%	2016-21	Ramm et al. (2018), Drakes et al. (2021)	0.84	0.09	0.13	0.81	0.86	0.91	0.97
Median Building Age	MedBLDGAge	Years	2016-21	Penning-Rowsell and Wilson (2006), Koc and Işık (2021)	41	12	13	34	43	46	101

4.4.2 Control Variables

Apart from the flood risk factors, this research has considered several other control variables. These variables are included in the causal model to control their effects so that true empirical relationships between the flood risk factors and the response variable can be derived. Table 4.1 presents the descriptive statistics of all the predictors including the flood risk factors. It also includes the existing literature that supports the inclusion of the chosen predictors.

1. **Annual Flood Damage**: It is evident that the extent of the flood damage influences the extent of flood insurance payout. The damage data was collected from the National Oceanic and Atmospheric Administration's (NOAA) Storm Events database. The database mentions the dollar amount of property damage for different flood and storm events. For this research, the flood damage in a county in a year was calculated as the sum of damage from flash floods, floods, and coastal floods in that county during that year.
2. **Annual Anomalous Rainfall**: The amount of rainfall is also an indicator for the level of flood damage. The rainfall data was collected from NOAA's website. In this research, the annual anomalous rainfall amount has been utilized as the predictor instead of the annual rainfall amount as extra rainfall is more relevant to flood risk than the actual amount. The anomaly is calculated as the difference of the annual rainfall and the mean annual rainfall between 1901 and 2000.
3. **NFIP Policies**: It is evident that the number of NFIP policies in a county and the total insured values (TIV) of those policies directly influence the expected NFIP payout in that county. Therefore, they were considered as the control variables for developing the regression model. It should be noted that the flood insurance data was only available for five years between 2016 and 2020. Hence, it was assumed that the number of NFIP policies and the total insured values of those policies remained the same as 2020 in 2021.
4. **County Characteristics**: A number of county characteristics were also considered in the causal model as control variables. They are population density, median building value, percentage of buildings that were occupied, and median building age.

Next, the pairwise correlations between the predictors were tested to ensure that there is no multicollinearity, i.e., high correlation among the predictors. Figure 4.5 shows the correlation matrix, where strong pairwise correlations can be noticed between (1) flood exposure and number of NFIP policies (0.78), (2) flood exposure and total insured value of the NFIP policies (0.84), (3) social vulnerability and community resilience (-0.49), and (4) number of NFIP policies and total insured value of the NFIP policies (0.97). To avoid multicollinearity, community resilience, number of NFIP polices, and total insured value of NFIP polices were removed from the list of predictors. After removing those three predictors, pairwise correlations between the remaining predictors were checked again. None of the correlation coefficient was more than 0.4 or less than -0.3, which suggests eradication of multicollinearity among the predictors.

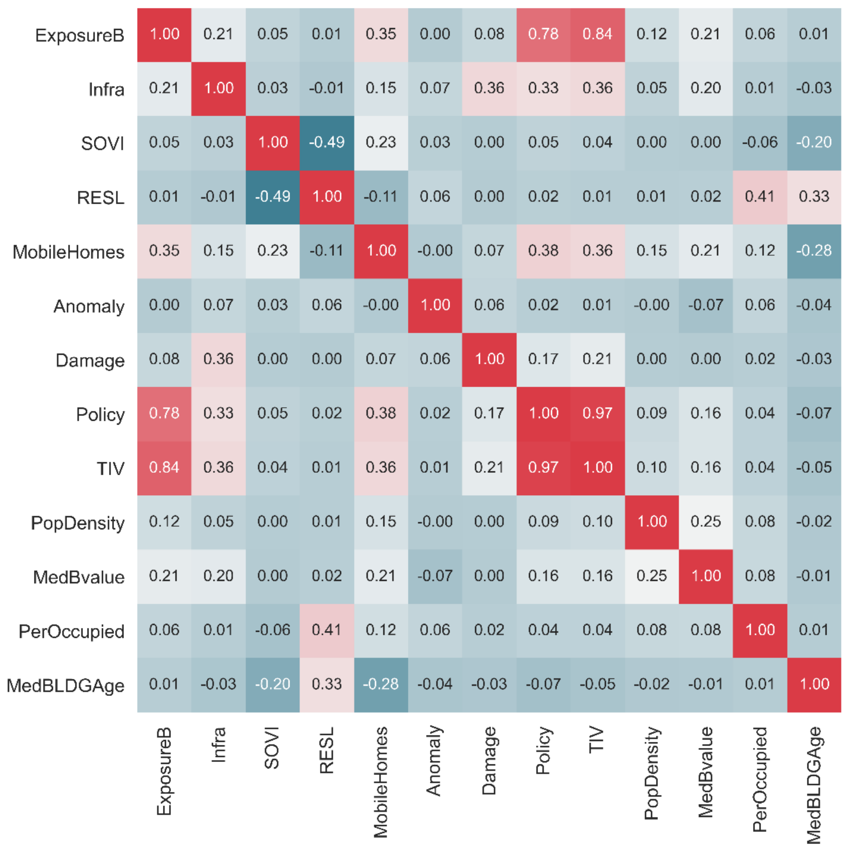


Figure 4.5 Correlation Matrix

Next, the predictors were standardized. This step was required to make sure that no variable is given higher or lesser significance due to the difference in their ranges. More importantly standardized regression coefficients can be compared to gauge the relative importance of different flood risk factors. Standardization follows equation 4.1 where x is a predictor, μ is the mean of the predictor, σ is the standard deviation of the predictor, and z is the predictor after standardization.

It should be noted that the standardized predictor, i.e., z has a mean of 0 and a standard deviation of 1. All values of z ranged between [-1, 1].

$$z = \frac{x - \mu}{\sigma} \tag{4.1}$$

4.4.3 Research Methods

In statistical literature, two types of models co-exist (1) inferential models that are used for causal explanations and (2) predictive models that are used for forecasting (Breiman 2001, Shmueli et al. 2010, Emmert-Streib and Dehmer 2021). It is important to note that the objective of this research is to develop a causal model therefore, prediction accuracy has not been prioritized. For developing the causal model, this research has adopted Linear Mixed Effects regression, which is also known as Linear Mixed Model (LMM). LMM is an extension of simple linear regression that can account for the hierarchical structure in the dataset, which for this research originates from data samples coming from different counties. LMM model is a popular choice in causal analysis and has been used in Huff et al. (2015), Bauer et al. (2018), Efendić (2021), etc. As explained earlier, the final dataset contained 4327 datapoints that belonged to 1813 unique counties. If one regression model with fixed regression coefficients is fitted to the whole dataset, it will not account for the heterogeneity among the samples. On the other hand, one regression model for each county is also not feasible due to lack of adequate datapoints. Therefore, the possible solution is to have fixed regression coefficients for the entire dataset along with random intercepts for different counties. The fixed coefficients account for the fixed effects part whereas the random intercepts account for the random effects part. Since it is a combination of fixed effects and random effects, it is known as a mixed effects model.

In addition to having random effects for each county, this research has considered the year (2016 to 2021) as a categorical variable to account for events that are common to all the counties in the dataset in a given year (Kousky et al. 2018). Equation 4.2 represents the LMM model fit by a restricted maximum likelihood estimation with a random effect for each county (1813 groups).

$$y = \alpha + \lambda_t + X\beta + Zu + \varepsilon \tag{4.2}$$

Where y is a N×1 vector of the response variable, i.e., log transformed annual NFIP payout of a county, α is a N×1 vector of fixed intercepts, λ_t is a N×1 vector of the year fixed effects, X is a N×p matrix of p predictors, β is p×1 vector of the fixed regression coefficients, Z is N×qJ design matrix of q random effects and J groups, u is a $qJ \times 1$ vector of q random effects for J groups, and ε is a $N \times 1$ vector of the residuals that follow a normal distribution with zero mean and constant variance, N is the number of datapoints, i.e., 4327, and lastly, for one random intercept, q is 1 and J is the number of unique counties, i.e., 1813. The group size varied between 1 and 6.

4.5 Results and Discussions

Table 4.2 presents the outcomes of the LMM. It can be noticed that all the p-values in Table 4.2 are less than 0.05, i.e., a 5% significance level. This indicates that all the predictors that have been used in the causal model are significant in explaining the changes in the annual NFIP payout in a county. The year fixed effects are also significant. These coefficients adjust the intercept for the regression model. For instance, the intercept for the year 2017 will be 12.15 – 0.34, i.e., 11.81. Similarly, the intercept for the year 2018 is 12.15 – 1.02, i.e., 11.13. It can also be noticed that Table 4.2 does not contain the coefficient for the year 2016 as 6 years have a degrees of freedom of 6 – 1, i.e., 5 years. Since the coefficient of 2016 is not included in Table 4.2, the intercept for the year 2016 will be 12.15 itself.

The rainfall anomaly and the flood damage have positive regression coefficients. This means that an increase in these two variables would increase the annual flood insurance claims in a county. Higher rainfall anomaly indicates higher precipitation than the average. If the rainfall anomaly is higher, there is more chance of floodings. As a result, there will be more flood insurance claims. The same is true for flood damage. For higher flood damage, the flood insurance claims are also expected to be higher.

The population density of a county has a positive regression coefficient. This implies that an increase in population density would lead to more flood insurance claims. It is a known fact that higher population density intensifies the flood impact in a region (Diakakis 2014, Santos et al. 2018). There are multiple reasons for that. First, higher population density increases the exposure. As a result, the potential of the flood affecting more households increases. More importantly, population density is positively correlated with the imperviousness of the surface (Hicks and Woods 2000, Sheng and Wilson 2009). Densely populated urban areas also have a higher concentration of infrastructure. Water cannot penetrate impervious surfaces like roads, buildings, parking lots, etc. Thus, converting wetlands and natural landscapes into impervious surfaces reduces infiltration of rainfall water into the soil and increases surface runoff. As a result, the severity of flooding increases.

Table 4.2 Regression Coefficients for LMM

Variable	**Coefficient**	**Std. Error**	**z**	**P>|z|**
Intercept	12.15	0.09	138.86	0.00
Year = 2017	-0.34	0.12	-2.88	0.00
Year = 2018	-1.02	0.12	-8.75	0.00
Year = 2019	-0.73	0.11	-6.62	0.00
Year = 2020	-0.98	0.11	-8.90	0.00
Year = 2021	-0.57	0.11	-5.13	0.00
Rainfall Anomaly	0.51	0.03	15.97	0.00
Flood Damage	0.24	0.03	8.20	0.00
Population Density	0.07	0.04	1.98	0.05
Median Building value	-0.12	0.03	-3.59	0.00
Percentage Occupied Buildings	0.08	0.03	2.29	0.02
Median Building Age	-0.19	0.04	-5.17	0.00
Flood Exposure	0.38	0.04	9.16	0.00
Infrastructure Vulnerability	0.16	0.03	5.16	0.00
Social Vulnerability	0.07	0.03	2.11	0.04
No. of Mobile Homes	0.23	0.04	5.57	0.00
Group Var	0.38	0.03		

In the U.S., 80% of the population live in the urban areas. Also, majority of the large cities in the U.S. are located in the coastal areas and therefore are exposed to coastal floods. The same trend was noticed in the current dataset, where the counties that were exposed to both river and coastal floods had a mean population density of 1291 per square miles and a median population density of 276 per square miles. On the other hand, the mean population density of the counties that were exposed to only river floods was 287 per square miles and the median was 95 per square miles. The difference in the population density is apparent. It has been explained before that the severity of coastal floods is higher than that of river floods. Thus, urban areas that are exposed to coastal floods contribute more to flood insurance claims.

Negative regression coefficients can be noticed for median building value. This indicates that lower median building value leads to higher flood insurance payout. This is consistent with previous findings in Knighton et al. (2020). It could be due to the depressed value of homes in low-lying lands. Siders et al. (2019) noted that the risk from climate related hazards is expected to decrease property value in the hazard exposed area. On the other hand, Grigg (2019) has found that repetitive loss properties account for 25-30% of the NFIP claims. It is safe to assume that the repetitive loss properties are located closer to the waterbodies and therefore exposed to severe flood hazard, which can lead to the decrease in the home price. Moreover, the Biggert-Waters Act of 2012 and the Grimm-Waters Act of 2013 that recommended reforms in NFIP placed more economic burden of flood losses on the at-risk properties. This reform has led to a reduction of floodplain property values (Indaco et al. 2019). The percentage of occupied buildings in a county has a positive regression coefficient. This means that an increase in the percentage occupancy will lead to more flood insurance claims. This might be related to flood insurance enrollment. With an increase in the percentage occupancy of the buildings, more people are expected to purchase flood insurance. That might increase the flood insurance claims.

However, the age of the building has a negative regression coefficient. Therefore, an increase in the median building age in a county will decrease the flood insurance claims. This result contradicts findings from most of the past literature. New buildings are often built following new standards that are more robust in the face of flood hazards. Therefore, when exposed to flood hazard of similar severity, it is expected that a county with lower median building age will suffer lower flood losses and have lower flood insurance claim than that of a county with higher median building age. However, the used data suggests that counties with lower median age of buildings contribute more to flood insurance claims. There are a few possible explanations for this. First, new buildings are often constructed in the flood zones. For instance, Climate Central, a New Jersey based research group in the U.S., analyzed real estate organization – Zillow's data and found that new construction in the 10-year flood zones in the U.S. has increased since 2010 (Flavelle 2019). Additionally, 24 U.S. cities including New York, Tampa, and Virginia Beach, etc., have built at least 100 homes in the 10-year flood zones since 2009. Newly constructed homes in a 10-year flood zone have a higher chance of having flood insurance. This might be the reason why the regression coefficient for median building age is negative. Second, Wing et al. (2010) have analyzed historical flood insurance claims outside of the 100-year flood zone and found that homes that were built after 1980 have historically suffered more flood damages hence had more flood insurance claims than that of homes that were built before 1980. This could be another reason why the regression coefficient for median building age is negative. Third, it has been explained before that flood insurance is mandated for a home located within a 100-year flood zone in a NFIP participating community and has a federally backed mortgage. Older homes have a higher likelihood of paid off mortgages. As a result, the mandatory requirement for flood insurance might not exist for them. Therefore, they are free to discontinue the flood insurance once the mortgage is paid off. This is another possible reason why counties with older homes could have less flood insurance claims. Lastly, the depreciated value of old buildings can also contribute to lower flood insurance claims after getting damaged by floods. Residential property value depreciates over time. Since flood insurance coverage accounts for the property value, depreciated property can have less coverage and therefore less flood insurance claims.

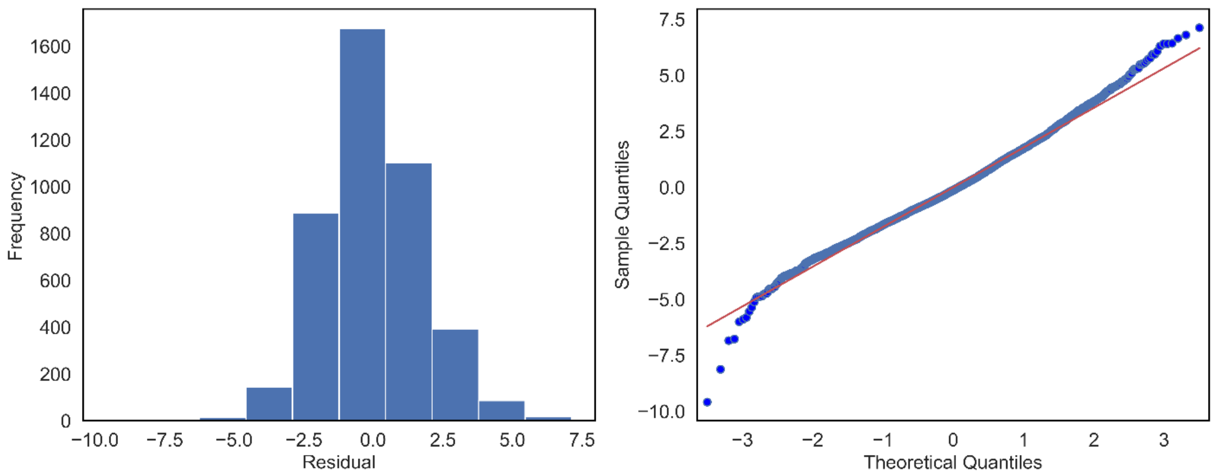
It has been explained in previous sections that all the flood risk factors considered in the causal model have reinforcing effects on flood damage and subsequently flood insurance claims. The outcomes of the causal model also support that notion. It can be noticed that all the flood risk factors have positive regression coefficients. Therefore, any increase in the flood risk factors will increase the flood insurance payout in the U.S. For instance, if flood exposure increases, more buildings will be exposed to flood hazard. Increased exposure will also increase the likelihood of flood damage and flood insurance claims. The corresponding regression coefficient estimates the increase in the flood insurance payout due to increase in flood exposure. It can be noticed that the regression coefficient has a value of 0.38. So, if the flood exposure in a county increases by one standard deviation, the natural logarithm of the annual flood insurance payout in that county will increase by 0.38 times of its standard deviation. So, the actual change in the flood insurance payout will be $e^{0.38}$, i.e., 1.46 times of its standard deviation. The standard deviations for the predictors and response variable can be noticed in Table 4.1. For instance, the standard deviation of flood exposure is \$21.7 billion whereas the standard deviation of the annual NFIP payout is \$71.8 million. Therefore, increasing the flood exposure by \$21.7 billion will increase the average annual NFIP payout by 1.46×\$71.8 million, i.e., \$104.8 million.

Similarly, vulnerable infrastructure also leads to higher flood insurance payouts. It has been explained before that infrastructure acts as the first level of resistance against natural hazards. As a result, vulnerable infrastructure exacerbates the severity of flooding, which increases the extent of flood insurance claims. The regression coefficient for infrastructure vulnerability is 0.16. Following the same process explained previously, it was found that if the infrastructure vulnerability, measured in terms of per capita public assistance payout, increases by one standard deviation, i.e., \$184, the average annual NFIP payout will increase by $e^{0.16}$×\$71.8 million, i.e., \$84.3 million. Social vulnerability decreases people's ability to attenuate the risk of natural hazards. Therefore, higher social vulnerability potentially leads to higher flood damage and subsequently higher flood insurance payout. If social vulnerability in a county that is quantified by CDC increases by one standard deviation, i.e., 0.27, the average annual NFIP payout will increase by $e^{0.07}$×\$71.8 million, i.e., \$77 million. Lastly, mobile homes are more likely to get damaged from natural hazards. If the number of mobile homes in a county increases by one standard deviation, i.e., 7704, the average annual NFIP payout will increase by $e^{0.23}$×\$71.8 million, i.e., \$90.4 million.

These relationships can help in forecasting future NFIP payout based on different counterfactuals. For instance, according to FEMA's National Risk Index, New Orleans – the largest city of Louisiana currently has a combined flood exposure of \$74.7 billion (\$45.5 billion from coastal flood and \$29.2 billion from river flood) (Zuzak et al. 2021). The 1st Street Foundation predicted that in 30 years the number of properties within the 100-year flood zone in New Orleans will become 1.67 times from 66,131 to 110,236 (1st Street Foundation 2023). Assuming that the coastal flood exposure will increase at the same rate, the combined flood exposure in New Orleans, Louisiana by 2030 will become \$124.5 billion, i.e., an increase by \$49.8 billion. Based on the developed causal model, this increase in flood exposure will increase the average annual NFIP payout by $\frac{(124.5-74.7)b\times 104.8m}{21.7b}$, \$241 million if everything else remains the same as present. It should be noted that this is a conservative estimate since it does not take into account the price inflation of homes in the calculation. The S&P CoreLogic Case-Shiller U.S. National Home Price NSA Index monitors the fluctuations in the value of the U.S. residential housing market by monitoring single-family home purchase prices. The S&P Dow Jones Indices LLC (2023) records show the index has increased by approximately 4 times from 76.4 in January 1993 to 305.1 in May 2023. If the increase in the housing price follows the same historical trend, it will become 4 times of the present price in the next 30 years. If the housing price increases by 4 times, the combined effect of housing price inflation and increased flood exposure by 67% will increase the expected annual NFIP payout by $\frac{(124.5\times 4-74.7)b\times 104.8m}{21.7b\times 1000}$, i.e., \$2.04 billion. The research results facilitate this type of counterfactual analysis. On the other hand, the relationships can be utilized to quantify the impact of a risk reduction strategy or policy on the annual NFIP payout. For instance, if property buyout is planned and estimated to reduce the flood exposure in New Orleans by 1%, it will reduce the expected annual NFIP payout by $20 million, i.e., 1% of $2.04 billion.

Since the flood risk factors have been standardized before developing the causal model, their regression coefficients can also be used to compare their relative importance for explaining the annual flood insurance payout. It can be noticed that flood exposure has the highest coefficient value among the flood risk factors, which is followed by mobile homes, infrastructure vulnerability, and social vulnerability. Therefore, among the flood risk factors, flood exposure is the most significant in explaining the annual flood insurance payout. It is followed by mobile homes, infrastructure vulnerability, and social vulnerability. This outcome is also important as it provides insight into which risk factor should be focused first for reducing the flood insurance payout. Clearly, reducing flood exposure should be given the maximum priority. Property buyout from high-risk flood zones can reduce flood exposure of a county (Yildirim and Demir 2021, Miao and Davlasheridze 2022). Hence, it can be a viable risk reduction strategy. However, Hegger et al. (2014) have suggested developing a diversified portfolio of flood risk reduction strategies that combine flood risk mitigation, adaptation, and recovery to maximize the benefits from the suite of strategies.

Lastly, the residuals from the regression model were checked to test the validity of the normality assumption of the residuals. Figure 4.6(a) shows the histogram of the residuals. The residuals are approximately symmetric with a skewness coefficient of 0.24, which is less than one. Figure 4.6(b) shows the Q-Q plot of the residual. In the Q-Q plot, the horizontal axis shows the quantiles of a standard normal distribution. The vertical axis shows the sample quantiles of the residuals from the regression model. The scatter plot appears to be on the straight line, which resembles a normal distribution. Therefore, it is safe to believe that the residuals of the regression model did not violate the normality assumption.



(a) (b)

Figure 4.6 Residual Plots. (a) Histogram (b) Q-Q Plot

4.6 Conclusion

In the last two decades a mixed methodology of statistical methods, data analytics, and machine learning techniques have become more prominent in disaster risk reduction studies. Researchers have been using historical data to develop various types of empirical models to derive insights from those data. Following a similar approach, this chapter has developed a causal model that can explain the annual flood insurance payout in a county based on different flood risk factors. To develop the model, historical data was collected between 2016 and 2021 and used in a mixed effects regression. The outcomes show that flood risk factors such as flood exposure, infrastructure vulnerability, social vulnerability, and mobile homes had a statistically significant reinforcing effect on the flood insurance payouts. Hence, an increase in the flood risk factors would increase the annual flood insurance payout. Among the flood risk factors, flood exposure had the highest impact on the flood insurance payouts. It has been estimated that if the flood exposure in a county increases by one standard deviation, i.e., \$21.7 billion, it will increase the average annual NFIP payout by \$104.8 million. Therefore, the federal government and FEMA should prioritize reducing flood exposure by implementing policies such as discouraging new constructions in floodplains, buying out at-risk properties, etc. On the other hand, the government should also discourage mobile home parks in the floodplain as they are more susceptible to flood damages. Improving the resilience of existing infrastructure and reducing social vulnerability can also help the federal government reduce the future cost of running the NFIP program. Since natural hazards are becoming more frequent and severe due to climate change, these risk reduction measures will be essential to keep the NFIP program solvent in future.

There are certain limitations to this research. First, the causal model has not been developed to predict the annual NFIP payout. Therefore, the predictive accuracy has not been tested and cross-validated. It should only be used for inferential purposes. In future, regression models that can accurately predict the future flood insurance payout can be developed using this data. Next, the flood exposure data used in the research was only available for one year. In absence of better alternative, it was assumed that flood exposure in a county was constant between 2016 and 2021, which might not be true due to climate change. Also, due to the dynamic nature of underlying control variables, the empirical relationships between flood risk factors and flood insurance payout might change in future. Therefore, the models should be updated periodically as more data becomes available. Attempts should also be made to develop the causal model at ZIP or postal code, census tract level. Also, the data driven models could be combined with existing hydrologic models to create more robust results.

5. PREDICTING ANNUAL FLOOD INSURANCE PAYOUT

Abstract

This chapter presents a regression model that can predict county level insured flood loss to households, measured in terms of annual flood insurance payout, in the U.S. based on different factors such as rainfall anomaly, flood damage, flood exposure, infrastructure vulnerability, social vulnerability, number of flood insurance policies, total insured value, etc. The regression model has been developed using five years of historical flood insurance claims data between 2016 and 2020. For developing the model, three regression techniques were adopted (1) Ordinary Least Square Regression, (2) Robust Regression, and (3) Generalized Linear Model. The final model, which uses the outcomes of those three models to predict the annual flood insurance payout for 2021, produced a coefficient of determination of 0.95 on the test set. The mean absolute error for predicting the annual flood insurance payout in a county on the test set was $831178. Overall, the total flood insurance claims across all the counties in 2021 was $1.68 billion. The proposed model predicted it as $1.85 billion, which is a 9.8% prediction error. Therefore, the proposed model could be used as a cheaper alternative for predicting the insured flood losses in the U.S.

5.1 Introduction

Natural hazards cause extensive damage to human life, infrastructure, property, economy, etc. Historical data from Munich Re shows that the frequency of natural hazards has increased steadily since 1980. During that period, disasters due to natural hazards have caused a cumulative loss of $5.2 trillion globally, the majority of which were uninsured (Munich Re 2020). Among the natural hazards, losses due to floods are by far the highest on a global scale (Colgan et al. 2017, Dubbelboer et al. 2017). The U.S. is no exception to that (Munich Re 2020).

There is no dearth of flood loss estimation models in the existing literature. For instance, Davenport et al. (2021) estimated that the cumulative loss from floods in the U.S. between 1988 and 2017 has been $199 billion. Quinn et al. (2019) have analyzed 40 years of historical flood data and found that there is a 1% chance that the losses from river floods would exceed $78 billion and a 0.1% chance of exceeding $136 billion in any given year in the U.S. Armal et al. (2020) have stated that the direct flood losses in the U.S. have risen from $4 billion annually in 1980 to $17 billion annually between 2010 and 2018. Jevrejeva et al. (2018) have forecasted that the global flood loss can exceed an additional \$1.4 trillion annually if the rise of global temperature is not maintained at 1.5 °C and reaches 2 °C. This can potentially cause an increase of the global sea level by an additional 11 cm.

Like other natural hazards, flood losses are also shared among the stakeholders. Peng et al. (2014) and Wang et al. (2020) have listed four classes of stakeholders who are associated with losses from natural hazards. They are households, primary insurers, reinsurers, and governments. However, the flood loss is not necessarily shared equally among the stakeholders. There are several factors that influence this cost sharing such as government's policy regarding flood insurance, insurance penetration rate, risk transfer to reinsurance, etc. Therefore, a stakeholder-centric flood loss and flood risk assessment is more insightful than a generic one as it reflects the true cost of floods to each class of stakeholder. Although there is plethora of flood loss and flood risk models, analysis on flood loss and flood risk from the perspective of a stakeholder is relatively underexplored, which emphasizes the need for this research.

In the U.S., the primary flood insurer and the government are the same entity. The government sponsored National Flood Insurance Program (NFIP) accounted for more than 95% of the primary residential flood insurance policies in 2018 (Kousky 2018). However, the flood insurance take-up rate in the U.S. is low. As a result, it has been observed in the aftermath of the flood related disasters that most of the sufferers are uninsured or underinsured (Kousky 2011). A congressional research report published in 2019 (Horn 2019) recorded the flood insurance take up rate for some of the recent flood events. The report estimated the average NFIP take up rate in the 100-year flood zones for multiple flood events. For instance, the NFIP take-up rate within 100-year flood zones after the South Carolina Flood in 2015 was found 30%, after Louisiana Flood in 2016 was found 31%, after Hurricane Harvey in Texas was found 21%, and after Hurricane Irma in Florida was found 31%. Additionally, it has been found that the NFIP policies are often short lived (Michel-Kerjan et al. 2012).

Under this situation, the federal government as the insurer of last resort compensates the disaster survivors who are underinsured and uninsured through FEMA managed Individual Assistance (IA) program. The Individuals and Households Program (IHP) within IA is the primary way the U.S. Federal Emergency Management Agency (FEMA) supports disaster survivors (Webster 2019). These two programs, i.e., NFIP and IHP together reflect the majority of the cost of floods to the U.S. federal government. Among the two, NFIP reflects the insured loss while IHP reflects the uninsured loss. To estimate the uninsured flood losses, it is essential to estimate the insured losses first. That insured flood loss can be measured in terms of the NFIP payout. Therefore, estimating NFIP payout can help in estimating the cost of floods to the federal government. However, to the best of the authors' knowledge, there is no existing model that can predict the extent of flood insurance claim, which is the gap this research aims to bridge.

This paper presents a prediction model that can predict the annual NFIP payout in a county based on different factors such as flood damage, flood exposure, infrastructure vulnerability, social vulnerability, community resilience, number of NFIP policies, number of NFIP claims, total insured value, etc. The model has been developed using five years of historical NFIP payout data between 2016 and 2020 by adopting three regression techniques (1) Ordinary Least Square Regression, (2) Robust Regression, and (3) Generalized Linear Model. The fourth and the final ensemble model uses the outcomes of these three regression models to generate its own predictions. The ensemble model was used to predict the annual flood insurance payout of the flood affected counties for the year 2021. The models produced a coefficient of determination of 0.95 and mean absolute error of $831178 on the test set. The proposed model estimates the extent of financial burden from flood loss on the primary insurer, which is also a significant part of the financial burden on the government, to help the Federal Emergency Management Agency (FEMA) in its financial preparedness to pay for the post-disaster flood insurance claims. Additionally, the proposed model is cheaper as it uses historical payout data in estimating the insured flood losses in future.

5.2 Research Background

The NFIP, which is managed by FEMA, was created in 1968 under the National Flood Insurance Act. The reluctance of private insurers to provide flood insurance created the need for the National Flood Insurance Program (Kousky et al. 2020). Purchasing flood insurance was voluntary till 1973. After 1973, buying flood insurance was mandated for properties with mortgage from a federally regulated or backed lender that are located within 100-year flood zone by the Flood Disaster Protection Act of 1973. Munich Re reported in 2020 that there were 14.6 million properties in the U.S. that were at substantial flood risk, i.e., located in 100-year flood zone. The report also stated that only 5% of all the single-family homeowners in the U.S. had flood insurance in 2020 (Munich Re 2020).

The NFIP has been considered problematic almost since its inception. Schilling et al. (1987) found that the program had been largely unsuccessful in the coastal areas. Historical records show that between 1978 and 2017, NFIP collected \$60 billion in premiums while paid \$65 billion as payouts (Grigg 2019). Payment of flood insurance claims is one of the key operating expenses of the NFIP. However, the program pays one-third of its income from premiums to the financial intermediaries to underwrite the flood insurance policies although none of the flood risk is borne by these intermediaries (Grigg 2019). Repetitive loss properties account for 25-30% of the claims although they are only about 1% of the insured properties (Grigg 2019). NFIP has been using flood maps for determining the flood insurance premiums for households and businesses. The U.S. Department of Homeland Security (DHS) reported in 2017 that only 42% of those flood maps can adequately identify flood risk (DHS 2017). Despite the low take up rate, NFIP had a cumulative debt of \$20.5 billion to the U.S. Treasury as of 2020 after the federal government forgave \$16 billion debt in 2016 (Grigg 2019, Horn 2020). It was projected in 2017 that the deficit of collected premium and the expected payout would remain \$1.4 billion in future (CBO 2017).

Researchers have long been recommending reforms in NFIP. Michel-Kerjan and Kunreuther (2011), Akabas (2014), McShane and Wie (2019) have recommended risk-based premiums that reflect the exposure of the homeowners, unsubsidized rates, protection of low-income groups, forgiveness of the debt to the U.S. treasury, reduction of exposure by reinsurance and CAT bonds, etc., for reforming NFIP. However, affordability of NFIP policies remains one of the key challenges for the lower-income households (Shively 2017). Frazier et al. (2020) claimed that an NFIP reform without any considerations for the socioeconomic vulnerability will create barriers for lower-income residents. On the other hand, Wagner (2022) has found that the willingness to pay for flood insurance is remarkably low in the U.S.

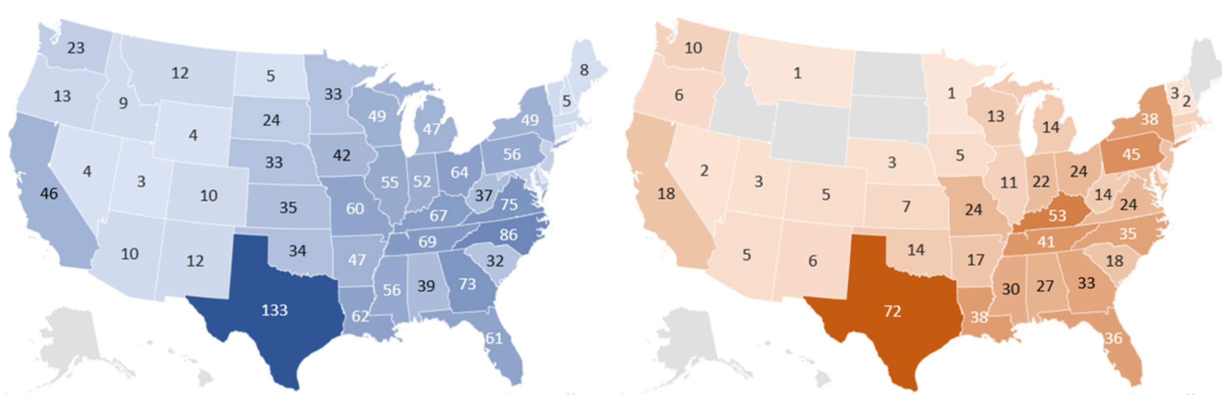
In 2021, FEMA brought a new actuarial approach of risk rating that relates the flood insurance premium of a household or business to its true flood exposure. Previously FEMA used flood maps, which did not always reflect the true flood risk of a property (DHS 2017). All new policies from October 1, 2021, onwards are subject to this new risk rating method. FEMA sees this new approach as a transformational leap forward that will make flood insurance premium rates fairer and more equitable. FEMA estimated that Risk Rating 2.0 will immediately decrease the monthly flood insurance premium for 23% of the policy holders. While 66% of them will see their monthly flood insurance premium increase by less than $10. The remaining 11% will face a monthly increase of flood insurance premium by more than $10 (FEMA 2022). However, there has been some controversy regarding this Risk Rating 2.0, as some states such as Louisiana could see an increase in premium for more than 80% of the policies (Murphy 2022).

Other than providing flood insurance to households, NFIP has a long-term objective to reduce federal expenditures on post-disaster assistance (Horn and Webel 2021). Therefore, NFIP is expected to ensure that future payouts are kept within limits. This requires the development of a model that can predict future payouts based on various factors. Utilizing historical payout data to derive empirical relationships between flood risk factors and NFIP payout can be one of the possible ways to enhance the understanding of NFIP payout. In the last two decades a mixed methodology of statistical methods, data analytics, and machine learning techniques have become more prominent in the flood risk related studies (Spekkers et al. 2014, Sadler et al. 2018, Desai and Ouarda 2021). Researchers have been using historical data to develop various types of empirical models to derive insights from those data.

Wing et al. (2020) have used historical NFIP redacted claims data to derive several insights on flood depth-damage functions that express flood damage to homes based on the depth of flood water inside the inundated home. They have found that the observed flood loss is a non-monotonic function of flood water depth. Mobley et al. (2021) have utilized the NFIP redacted claims data to develop a continuous flood hazard map for the Texas Gulf Coast region using a Random Forest model. Knighton et al. (2020) have also used historical flood insurance claims data to develop a model that can predict the number of parcel-level and census tract-level flood insurance claims in New York state using Random Forest classification and regression model. While their model is useful, it was only developed for the New York state. More recently, Ghaedi et al. (2022) have also utilized NFIP redacted claims data to predict the extent of flood losses in terms of the number of flood insurance claims based on different factors such as flood peak ratio, Giovanni flooded fraction, land slope, population per area, number of NFIP policies, etc. While the number of flood insurance claims is an important proxy for flood damage, it does not reflect the complete picture. The extent of flood damage is better reflected in terms of the total payout as it shows the magnitude of the financial burden. Additionally, their research did not consider factors such as infrastructure vulnerability, social vulnerability, existing resilience, etc. These factors influence the extent of flood damage (Choi et al. 2019, Sanders et al. 2020, Koc and Işık 2021), which influences the NFIP payout. Therefore, the reviewed literature suggests that there is a need to develop a data-driven multivariate prediction model that can predict the expected annual flood insurance payout so that it can be used by the decision makers to predict insured flood losses in future.

5.3 Research Data and Methods

This research has collected data from multiple sources for six years between 2016 and 2021. The regression models were developed using five years of data between 2016 and 2020, and the model performance was tested on 2021 data. All the collected data was aggregated at the county level to develop an unbalanced panel dataset where the panel unit is the county. It is important to note that the research has only considered the counties that had an annual flood insurance claim greater than zero and less than $500 million from 2016 to 2020 for training the models. In the training set, there were 1709 unique counties. On the other hand, there were 783 unique counties in the test set. Figure 5.1 shows the spatial distribution of the counties in the training set (Figure 5.1.a) and test set (Figure 5.1.b). It can be observed that the training set has representation from 48 states that excluded Alaska and Hawaii. Also, majority of the counties belong to the eastern half of the U.S. On the other hand, the test set is comprised of counties from 43 states, which excluded Alaska, Hawaii, Idaho, Maine, North Dakota, South Dakota, and Wyoming.



(a) (b)

Figure 5.1 Spatial Distribution of Counties Analyzed in this research.
(a) in Training Set and (b) in Test Set

The response variable for the prediction model is the annual NFIP payout in a county. The data was collected from FEMA's open data portal that has the records of all the redacted NFIP claims. The annual NFIP payout in a county varied between zero and \$4.4 billion in the training set. However, for developing the regression models, only the counties that had an annual NFIP payout of more than zero but less than \$500 million were considered. This step removed 4 counties from the training set, which is approximately 0.11% of the total number of the remaining datapoints in the training set. The final training set had 3542 datapoints, each of which represented a county. There were 574 unique counties in 2016, 581 unique counties in 2017, 723 unique counties in 2018, 860 unique counties in 2019, and 803 unique counties in 2020. The descriptive statistics of the response variable can be found in Table 5.1.

5.3.1 Predictor Variables

For predicting the annual NFIP payout in a county, the regression models used different predictors. Table 5.1 lists all of them along with their descriptive statistics. It should be noted that the descriptive statistics are calculated on the training set only. Table 5.1 also demonstrates the literature that supports the inclusion of predictors. It is important to note that the supporting literature only claims relationships between the predictors and the flood risk in an area. Since the existing flood risk in an area influences the flood insurance payout in that area, it is safe to assume that the predictors influence the flood insurance payout, i.e., NFIP payout. The predictors were categorized into four types. They are (1) County Characteristics, (2) Hazard Characteristics, (3) Flood Risk Factors, and (4) Flood Insurance Characteristics.

The first type of predictors is the county characteristics. These variables are the confounding factors that influence both the predictors and the response variables. The county characteristics that were considered in this research are population, area, median building value, percentage of occupied buildings, median building age, and climate zone. The first five county characteristics data was collected from the U.S. Census Bureau. In addition to these five county characteristics, this research has considered the climate zone of a county as a categorical predictor. The National Centers for Environmental Information identified nine climate zones in the contiguous U.S. (Karl and Koss 1984). Through this variable, the climate variation among the counties were captured. The percentages of counties in climate zones one to nine in the training set were 11%, 10%, 24%, 21%, 5%, 21%, 2%, 3%, and 3%, respectively. On the other hand, the percentages of counties in climate zones one to nine in the test set were 19%, 4%, 24%, 22%, 1%, 23%, 2%, 2%, and 3%, respectively.

The hazard characteristics included rainfall data, flood damage data, and type of flood risk (river and/or coastal). It is evident that the extent of the flood damage influences the extent of flood insurance payout. The damage data was collected from the National Oceanic and Atmospheric Administration's (NOAA) Storm Events database. The database mentions the dollar amount of property damage for different flood and storm events. For this research, the flood damage in a county in a year was calculated as the sum of damage from flash floods, floods, and coastal floods in that county during that year. The amount of rainfall is another indicator for the level of flood damage. The rainfall data was collected from NOAA's website. In this research, the annual anomalous rainfall amount has been utilized as the predictor instead of the annual rainfall amount as extra rainfall is more relevant to flood risk than the actual amount. The anomaly is calculated as the difference of the annual rainfall and the mean annual rainfall between 1901 and 2000. The type of flood risk is another categorical variable that can have two possible values, i.e., river flood risk and/or coastal flood risk. Among the counties in the training set, 81% were exposed to river floods whereas the remaining 19% were exposed to river and coastal floods. On the test set 70% of the counties were exposed to river flood risk whereas the remaining 30% were exposed to river and coastal flood risk.

Table 5.1 List of Variables with Descriptive Statistics on Training Set

Variable	Variable Type	Variable Code	Unit	Data	Supporting Literature	Mean	SD	Min	25%	50%	75%	Max
Annual NFIP Payout	Response	NFIP	$	2016-21	NA	2278279	18326080	8	26849	87233	357421	428323100
Population	Predictor	Population	Count	2016-21	Rasch (2016), Moreira et al. (2021)	245470	581259	1296	28417	68209	222277	10105720
Area	Predictor	Area	Sq. Miles	2016-21	Ouma and Tateishi (2014), Moreira et al. (2021)	808	1077	2	427	603	864	20057
Median Building Value	Predictor	MedBValue	$	2016-21	Schröter et al. (2014), Wing et al. (2020)	4498310000	29615950000	39500	118600	163800	280600	951000000000
Percentage of Occupied Buildings	Predictor	PerOccupied	%	2016-21	Ramm et al. (2018), Drakes et al. (2021)	0.84	0.09	0.13	0.81	0.86	0.91	0.97
Median Building Age	Predictor	MedBLDG Age	Years	2016-21	Penning-Rowsell and Wilson (2006), Koc and Işık (2021)	40	12	13	33	43	46	101
Climate Zone	Predictor	Region	Categorical	Fixed	Peng et al. (2022)	NA	NA	NA	NA	NA	NA	NA
Rainfall	Predictor	Anomaly	Inches	2016-21	Tarhule (2005), Zhang et al. (2018)	8.69	8.42	-18.88	2.68	8.18	14.64	39.41

Table 5.1 continued

Flood Damage	Predictor	Damage	$	2016-21	Botzen et al. (2009), Boamah et al. (2015)	9612187	191111100	0	0	0	113000	8000001000
Type of Flood Risk (River and/or Coastal)	Predictor	Risk	Categorical	Fixed	Haer et al. (2018)	NA	NA	NA	NA	NA	NA	NA
Flood Exposure	Predictor	ExposureB	$	2022	Stephenson and D'ayala (2014), FEMA (2021)	5679103000	21616840000	0	193725100	524177800	2412027000	313000000000
Infrastructure Vulnerability	Predictor	Infra	$	2016-21	Len et al. (2018), Sanders et al. (2020)	1449542	8066068	0	0	0	397339	338415500
Mobile Homes	Predictor	Mobile Homes	Count	2016-21	Baker et al. 2014, Rumbach et al. 2020	4990	7681	0	1372	2797	5518	89895
Social Vulnerability	Predictor	SOVI	NA	2016, 2018, and 2020	Cutter et al. 2003; Zhang and You 2014	0.55	0.27	0.00	0.32	0.56	0.78	1.00
Community Resilience	Predictor	RESL	NA	2020	Cutter et al. (2014), Choi et al. (2019)	55.12	2.58	45.98	53.40	55.21	56.99	64.67
No. of NFIP Policies	Predictor	Policy	Count	2016-20	Owusu-Ansah et al. (2019), Moreira et al. (2021)	4377	16394	5	146	462	1915	324898
Total Insured Value	Predictor	TIV	$	2016-20	Patankar and Patwardhan (2016), Wang and Sebastian (2021)	1845072000	7471702000	811928	34589860	141751400	728248200	129000000000

The flood risk factors included flood exposure, infrastructure vulnerability, mobile homes, social vulnerability, and community resilience. The flood exposure data was collected from FEMA's National Risk Index (Zuzak et al. 2021). To estimate the exposure to river and coastal flood hazards, the hazard occurrence and susceptible zone polygons (as suggested by FEMA) are overlapped with the appropriate administrative polygons (county for this paper). The resulting intersecting shape measures the area of exposure. The number of buildings within that area is the measure of flood exposed buildings and the aggregated property value of those flood exposed buildings represents the flood exposure. The detailed procedure for the exposure calculation can be found from FEMA's National Risk Index's Technical Documentation (Zuzak et al. 2021). The flood exposure in a county is calculated as the sum of the exposure from river and coastal flood hazard. It should be noted that the data on flood exposure is only available for one year. The values are in 2022-dollars. Therefore, the exposure has been assumed constants for all six years between 2016 and 2021 due to lack of better data. Another factor that increases the likelihood of flood damage is the number of mobile homes in a county. It has been found that mobile homes are more prone to flood damage (Baker et al. 2014, Rumbach et al. 2020). Hence, it has been considered as a predictor in this research.

Infrastructure works as the first line of protection against natural hazards and vulnerability is defined as the measure of proneness to threat scenarios that include natural hazards, intentional attacks, etc., (Ezell 2007). Vulnerable infrastructure often leads to more damage from natural hazards. U.S. federal government reimburses the state, local, tribal, and territorial (SLTT) governments to repair, restore, reconstruct, or replace their disaster damaged infrastructure in case of a presidentially declared major disaster through its Public Assistance (PA) program (Congressional Research Service 2021). The amount of federal support received by a county can be used as proxy variable to gauge the infrastructure vulnerability in that county (Bhattacharyya et al. 2023). The PA data was collected from FEMA's open data portal. The PA dataset was used to calculate the annual public assistance payout in county, which has been used as the predictor in the prediction model.

Social vulnerability includes the socio-economic and demographic factors that increase or decrease the impacts of natural hazards on a community (Tierney et al. 2001, Heinz Center 2002). Previous researchers have found that the impacts of disasters are biased towards the socially vulnerable population. For instance, Campbell et al. (2020) have found that vulnerable populations suffer the most damage from floods. These people include seniors, people with functional and access needs, people of lower economic status, and other minorities. Several research works have analyzed the role of multiple socio-demographic factors on flood vulnerability and have found significant impact (Cutter et al. 2003, Zhang and You 2014, Dandapat and Panda 2017, Emrich et al. 2020, Drakes et al. 2021, Koc and Işık 2021). This research utilizes the Social Vulnerability Index (SVI) proposed by the Centers for Disease Control and Prevention (CDC). The index is developed by utilizing 16 socioeconomic variables such as poverty level, unemployment, health insurance, race, ethnicity, disability, elderly population, etc. The SVI data for each county, which was collected from CDC's website, was only available for the years 2016, 2018, and 2020. Therefore, the SVI for 2017 and 2019 were calculated by interpolating, i.e., by taking average in this case. Additionally, when this research was conducted, the SVI for 2021 was not published. So, the SVI for 2021 was assumed to be equal to that of 2020.

The National Institute of Standards and Technology (NIST) has defined Community Resilience as the ability of a community to prepare for anticipated natural hazards, adapt to the altering conditions, and withstand and recover speedily after the disaster (NIST 2020). The existing coping capacity of a community is an important predictor for estimating the impact of a natural hazard in that community (Scheuer et al. 2011, Yang et al. 2013, Terti et al. 2015). Choi et al. (2019) have proposed that a disaster resilient community needs capacities in its all seven layers of critical infrastructures. These seven layers are civil, civic, social, educational, financial, environmental, and cyber. However, this research has utilized the Baseline Resilience Indicator for Communities (BRIC) developed by University of South Carolina's Hazards and Vulnerability Research Institute because of the availability of relevant data (Cutter et al. 2014). The indicator is developed based on 49 factors representing six types of resilience: social, economic, community capital, institutional capital, housing or infrastructure, and environmental (Cutter et al. 2014). The community resilience data was also collected from FEMA's National Risk Index database for the year 2020. The community resilience has been assumed constant for other years due to lack of available data.

The fourth type of predictors are related to the flood insurance characteristics of the county. It is evident that the number of NFIP policies in a county and the total insured values (TIV) of those policies directly influence the expected NFIP payout in that county. Therefore, they were considered as the predictors for developing regression models. It should be noted that the flood insurance data was only available for five years between 2016 and 2020. Hence, it was assumed that the number of NFIP policies and the total insured values of those policies remained the same as 2020 in 2021. Next, pairwise correlations between the predictors were checked. Figure 5.2 shows the correlation matrix.

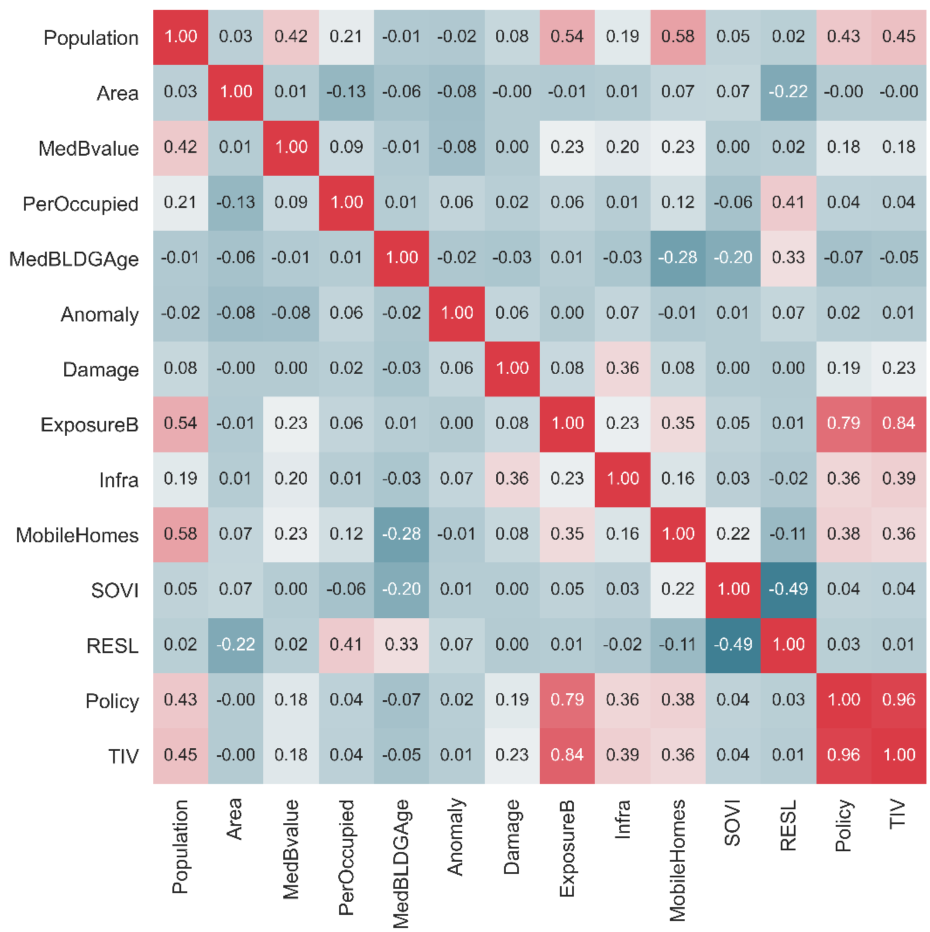


Figure 5.2 Correlation Matrix

It is important to note that the correlation matrix does not contain the two categorical predictors as they are not continuous variables. From Figure 5.2, strong pairwise correlations can be noticed between (1) population and flood exposure (0.54), (2) population and mobile homes (0.58), (3) flood exposure and number of NFIP policies (0.79), (4) flood exposure and total insured value of the NFIP policies (0.84), (5) number of NFIP policies and total insured value of the NFIP policies (0.96), and (6) social vulnerability and community resilience (-0.49). To avoid multicollinearity, population, community resilience, number of NFIP polices, and total insured value of NFIP polices were removed from the list of predictors. After removing those three predictors, pairwise correlations between the remaining predictors were checked again. None of the correlation coefficient was more than 0.4 or less than -0.3, which suggests eradication of multicollinearity among the predictors.

5.3.2 Research Methods

The research methodology is shown in Figure 5.3. The first step involved data cleaning, aggregating, testing pairwise correlations, etc. Next, four prediction models were developed using all the predictors presented in Table 5.1 excluding population, community resilience, number of NFIP policies, and total insured value to avoid any multicollinearity issue. In statistical literature, two types of models co-exist (1) inferential models that are used for causal explanations and (2) predictive models that are used for forecasting (Breiman 2001, Shmueli et al. 2010, Emmert-Streib and Dehmer 2021). It is important to reiterate that the objective of this research is to develop prediction models that can predict the response variable, i.e., annual NFIP payout in a county with adequate accuracy. Therefore, the models should only be used for prediction purposes and not for inferential, i.e., causal purposes.

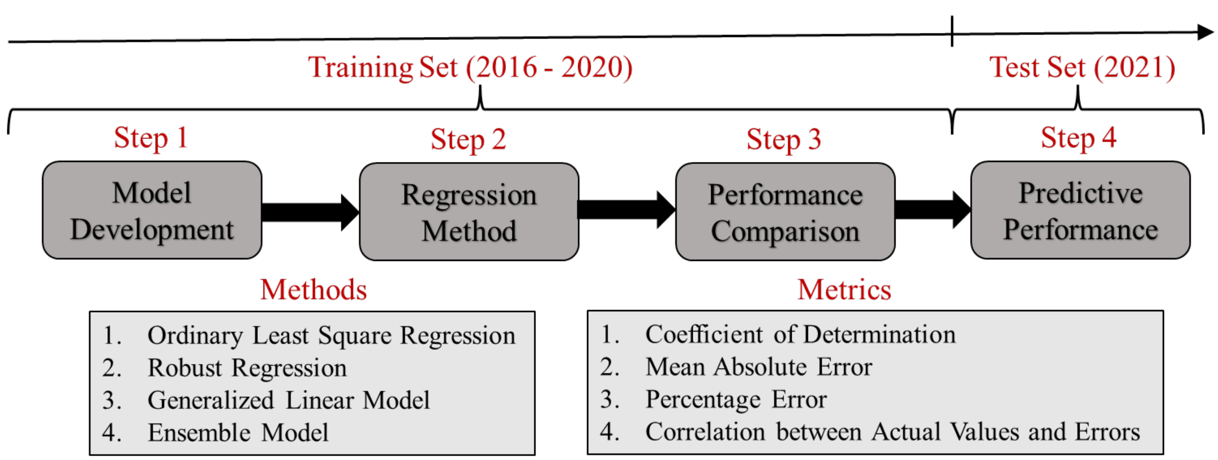


Figure 5.3 Research Framework

For developing the models, three different regression techniques have been used. They are (1) Ordinary Least Square Regression, (2) Robust Regression, and (3) Generalized Linear Model. The final ensemble model used the outcomes of these three models to derive its own predictions. These regression models are simpler to develop and understand. Also, they provide the flexibility of analyzing the coefficients of different predictors to understand how a predictor is influencing the response variable. More importantly, these models produced adequate prediction accuracy for this research. Hence, this research did not try any other advanced models.

As explained earlier, the models were developed using five years of data between 2016 and 2020. The models' performances, i.e., goodness of fit was quantified based on four metrics (1) the coefficient of determination (R^2), (2) mean absolute error (MAE), (3) percentage error, and (4) correlation between the actual values of the response variables and the corresponding prediction errors. The percentage error measures the deviation of the predicted sum of NFIP payout of all counties and the actual sum of NFIP payout of all the counties. This variable is important to FEMA because it reflects the total amount of flood insurance claims that will generate in a year.

Ordinary Least Square Regression (OLS)

Linear regression is the simplest form of regression that assumes a linear relationship between the explanatory and response variable. It fits a straight line or surface that minimizes the differences between the actual and the predicted values of the response variable. The linear regression technique uses the least squares method to find the best fit line for a set of paired data. The line is represented by equation 5.1, where y is the response variable, i.e., the annual NFIP payout in county, *a* is the intercept, *b* is the vector of slopes of the line, *X* is the vector of predictors, and ϵ is the residual error that follows a normal distribution with zero mean and constant variance σ.

$$y = a + Xb + \epsilon \tag{5.1}$$

The ordinary least square method minimizes the mean squared error (MSE) shown in equation 5.2 where *N* is the number of datapoints to derive the values of *a* and *b*.

$$MSE = \frac{1}{N}\sum_{n=1}^{N}(y - (a + Xb))^2 \tag{5.2}$$

Robust Regression (RR)

Linear regression often suffers due to the presence of outliers in the dataset. The outliers are the points that have high residual errors. These points can have substantial influence of the regression coefficients. Robust regression is a substitute to ordinary least square regression when the data is dominated by outliers or influential observations. The regression equation is the same as the linear regression that is shown in equation 5.1. However, the objective function for robust regression is different from equation 5.2. The objective function for robust regression is shown in equation 5.3, where ρ is a function of the residual error ϵ. It can be noticed that for ordinary linear regression, $\rho(\epsilon_i) = \epsilon_i^2$ resembles the objective function of linear regression, i.e., equation 5.2.

$$\sum \rho(\epsilon_i) = \sum_{n=1}^{N} \rho(y - (a + Xb)) \tag{5.3}$$

The robust regression uses a method named Iteratively Reweighted Least Squares (IRLS) to assign weights to each datapoint. This method is less sensitive to major changes in small parts of the data thus making it less sensitive to outliers. The IRLS algorithm iteratively computes the weights of the datapoints. At initialization, the algorithm gives equal weight to each data point. Then it estimates the model coefficients using ordinary least squares, i.e., the method followed in ordinary linear regression. The algorithm computes the weights (w_i) after each iteration. It assigns lower weights to points that are farther from the model predictions in the previous iteration. This research has used the Huber estimator for the objective function and weights (Huber 1964). The Huber estimator computes the objective function and the weight function as equation 5.4 and 5.5, respectively, where k is the tuning constant and equals to 1.345 times of the standard deviation of the residual error, i.e., σ.

$$\rho_H(\epsilon) = \begin{cases} \frac{1}{2}\epsilon^2 & for\ |\epsilon| \le k \\ k|\epsilon| - \frac{1}{2}k^2 & for\ |\epsilon| > k \end{cases} \tag{5.4}$$

$$w_H(\epsilon) = \begin{cases} 1 & for\ |\epsilon| \le k \\ \frac{k}{|\epsilon|} & for\ |\epsilon| > k \end{cases} \tag{5.5}$$

The algorithm then computes model coefficients (a and b) using the weighted least squares method. Iteration is stopped when the coefficient values converge within a specified limit. This algorithm tries to find the curve that fits the majority of the data using the least-squares approach, while minimizing the effects of outliers.

Generalized Linear Model (GLM)

Linear regression assumes normal distribution of the residuals. This assumption does not always hold. In such cases, Generalized Linear Model (GLM) is adopted. In these models, the response variable is assumed to follow a distribution from the exponential family, which could be Poisson, Gamma, Tweedie, etc. For modeling the claims in the insurance industry, where a lot of datapoints are clustered in the region of zero, Tweedie distribution is popular (Shi 2016, Yang et al. 2018, Fontaine et al. 2020). Hence, it has been used in this research. There are three components of a GLM.

- Random Component: It denotes the probability distribution of the response variable. For this research, the response variable is assumed to follow a Tweedie distribution.
- Systematic Component: It denotes the linear combination of the predictor variables.
- Link Function: It denotes the link between the Random component and the Systematic component. The logarithmic function has been used as the link function for this research.

The mean regression structure of the Tweedie regression is shown in equation (5.6)

$$\log(\mu) = a + bX \tag{5.6}$$

The response variable $y \sim$ Tweedie $(\mu, \varphi\mu^p)$, where μ is mean, φ is the scale parameter ($\varphi > 0$), and p is the power parameter for the Tweedie distribution, which was kept at 1.05 for this research. The variance of the distribution is measured by $\varphi\mu^p$. The parameters are estimated through Maximum Likelihood Estimation.

5.4 Results

Table 5.2 presents the regression coefficients for the three regression models developed using multiple linear regression (OLS), robust regression (RR), and generalized linear model (GLM). Among the predictors, positive coefficients can be noticed for flood damage, rainfall anomaly, flood exposure, infrastructure vulnerability. This indicates that an increase in these variables would increase the annual NFIP payout in a county. Although the regression coefficients are important, the primary objective of this research is to predict the annual NFIP payout with adequate accuracy. Hence, it is more important to assess the predictive performance of the models.

Table 5.2 Regression Coefficients for Initial Models

	OLS		RR		GLM	
Predictors	**Coeff**	***p*-value**	**Coeff**	***p*-value**	**Coeff**	***p*-value**
Intercept	-5930000	0.048	-4023000	0.000	12.6	0.000
Flood Damage	0.05	0.000	0.05	0.000	5.03×10^{-10}	0.000
Rainfall Anomaly	146100	0.000	18400	0.000	0.06	0.000
Area	220	0.482	90	0.000	2.00×10^{-4}	0.127
Median BLDG Value	-8.54×10^{-6}	0.311	-4.68×10^{-6}	0.000	-1.70×10^{-11}	0.000
Percentage Occupied	1262000	0.622	173600	0.226	0.20	0.764
Median BLDG Age	-22960	0.363	-5751	0.000	-0.01	0.157
Flood Exposure	3.24×10^{-5}	0.010	2.85×10^{-5}	0.000	3.54×10^{-12}	0.008
Infra Vulnerability	0.65	0.000	0.48	0.000	2.52×10^{-8}	0.000
Social Vulnerability	-1889000	0.070	-251600	0.000	-1.01	0.000
Mobile Homes	128	0.001	31	0.000	4.18×10^{-5}	0.000
River Flood Risk	-4318000	0.000	-688500	0.000	-2.5	0.000
Climate Region 1	6955000	0.000	4200000	0.000	0.8	0.286
Climate Region 2	8774000	0.000	4531000	0.000	2.3	0.004
Climate Region 3	9211000	0.000	4725000	0.000	2.4	0.001
Climate Region 4	7573000	0.000	4499000	0.000	2.2	0.001
Climate Region 5	6921000	0.001	3798000	0.000	2.5	0.006
Climate Region 6	10970000	0.000	4940000	0.000	3.3	0.000
Climate Region 7	8166000	0.001	4145000	0.000	-0.6	0.744
Climate Region 8	7088000	0.001	4050000	0.000	0.1	0.950

As explained previously, the predictive performances of the models have been assessed based on (1) the coefficient of determination (R^2), (2) mean absolute error (MAE), (3) percentage error, and (4) correlation between the actual values of the response variables and the corresponding prediction errors. Table 5.3 displays the performance metrics for OLS, RR, GLM, and the Ensemble model on the training set. It has been explained before that the ensemble model used the outcomes of the OLS, RR, and GLM models to predict the annual NFIP payout in a county. The prediction of the ensemble model is calculated as the average of the predictions from the OLS, RR, and GLM models.

Table 5.3 Predictive Performance of Initial Models on the Training Set

Performance Metrics	OLS	RR	GLM	Ensemble
R^2	0.37	0.41	0.45	0.46
MAE	$2,973,688	$2,221,755	$2,617,809	$2,566,557
Percentage Error	24.89%	-33.96%	-0.04%	-3.05%
Error Correlation	0.76	0.83	0.67	0.77

From Table 5.3, it can be noticed that none of the four models has performed satisfactorily. In terms of the coefficient of determination, the ensemble model has performed the best, although the R^2 value is not high. The MAE shows the average error in predicting the annual NFIP payout of a county. It can be noticed that the robust regression model has the lowest MAE among the four models. This is due to the ability of robust regression to reduce weights of outliers in the dataset. However, the MAE is still very high. For each county, the RR model on average produced an error of $2.2 million. The percentage error shows the difference between the total actual NFIP payout and the total predicted NFIP payout as a percentage of the total actual NFIP payout. It can be noticed that the GLM and the ensemble modes have predicted the total payout with very high accuracy. However, both models struggle in terms of the other three performance metrics.

Due to lack of adequate predictive performance of the initial models, this research has considered an additional predictor, which is the number of annual NFIP claims in a county. As explained in the Research Background section, Ghaedi et al. (2022) have recently developed a multivariate prediction model that can predict the number of flood insurance claims in a county based on different flood characteristics with adequate accuracy. This research takes advantage of that development and uses it as a predictor for predicting the expected flood insurance payout. Therefore, the final model uses all the predictors from the initial models as shown in Table 5.2 along with the number of annual NFIP claims in a county for prediction purposes. Again OLS, RR, GLM, and ensemble models were developed following the same procedure. Like the initial models, the ensemble model in this case also derived its predictions based on the outcomes from the OLS, RR, and GLM models. Table 4 shows the predictive performance of the final models.

Table 5.4 Predictive Performance of Final Models on Training Set

Performance Metrics	OLS	RR	GLM	Ensemble
R^2	0.93	0.91	0.64	0.93
MAE	$842,985	$781,526	$1,988,887	$885,639
Percentage Error	11.27%	-1.61%	0.20%	3.29%
Error Correlation	0.28	0.60	-0.09	0.26

It can be noticed that the predictive performances have significantly improved after the addition of the new predictor. In terms of the coefficient of determination, the OLS, RR, and ensemble model have produced high values. The MAE values of the OLS model and the ensemble model are similar. The RR model has produced the lowest MAE, which is due to giving less weights to the outliers. In terms of the percentage error, the RR and ensemble model produced similar performances. However, the correlation between the error and the actual values is significantly higher in the RR model than that of the ensemble model. This indicates the RR model struggled with predicting the NFIP payout of those counties where the values are in the higher region. This is again due to giving less weights to those counties that have higher annual NFIP payouts. Based on all these performance metrics, this research chose the ensemble model for the prediction purpose.

Table 5.5 shows the performance of the final ensemble model on the test set. It can be noticed that the final ensemble model has produced a coefficient of determination of 0.95, which is very high. The MAE is $831178, which is better than the training set. However, the correlation between the actual values and the residual error is still high. This can be considered as one of the limitations of this model. The percentage error is also less than 10%. In 2021, the total actual NFIP payout was $1.68 billion whereas the final ensemble model predicted total annual NFIP payout of $1.85 billion. Therefore, the model overestimated the total flood insurance payout by approximately $170 million. Figure 5.4 shows the scatter plot of the actual vs predicted annual NFIP payouts of 783 counties in the test set. It can be noticed that the majority of the datapoints either fall on the regression line or very close to the regression line.

Table 5.5 Predictive Performance of Final Ensemble Model on Test Set

	R^2	MAE	Percentage Error	Error Correlation
Test Set	0.95	$831,178	9.79%	0.60
Null Model	NA	$3,581,886	NA	NA

Lastly, to prove the robustness of the predictions, the performance of the ensemble model was compared to that of the null, i.e., mean only model. In the null model all predictions are assumed to be equal to the mean of the variable. It is a popular benchmark for testing the statistical power of the prediction model in explaining the variance of the data. The performance of the null model is also shown in Table 5.5. It can be noticed that the MAE of the null model is reduced by approximately 77% by the final ensemble model, which proved the robustness of the developed model.

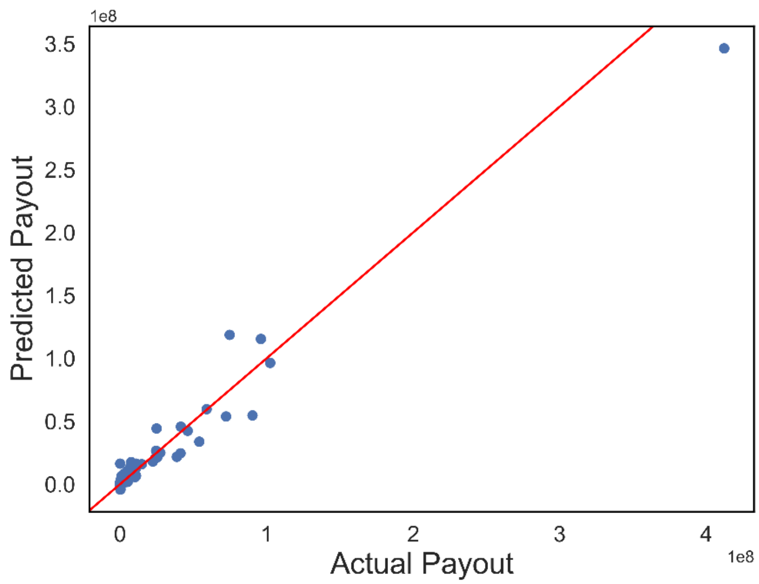


Figure 5.4 Scatter Plot of Actual vs Predicted Annual NFIP Payout on Test Set

In the next step, the nature of the residual error in each state was tested. Table 5.6 shows the actual and predicted NFIP payout for each state. It can be noticed that Louisiana received the highest NFIP payout in 2021. This is due to the widespread destruction caused by Hurricane Ida. It also caused severe floods in the northeastern states such as New York, New Jersey, Pennsylvania, and Connecticut (Livingston 2021). The actual NFIP payout in Louisiana in 2021 was approximately $959 million, which the ensemble model predicted as $969 million. The actual and predicted amount for New York were $111 million and $103 million, respectively. For New Jersey, the actual NFIP payout was $274 million, which was predicted as $247 million. It can be noticed that the model has overestimated the annual NFIP payout for multiple states such as Texas, Florida, Alabama, Mississippi, North Carolina, Connecticut, etc. This can be considered another limitation of the proposed model.

Table 5.6 State wise Comparison of Actual and Predicted Annual NFIP Payout in 2021

State	Actual NFIP Payout in 2021	Predicted NFIP Payout in 2021	Error in Prediction
AL	$20,305,460	$24,889,210	$4,583,750
AR	$3,030,171	$5,684,480	$2,654,309
AZ	$2,929,048	$3,928,291	$999,243
CA	$4,287,283	$4,602,869	$315,586
CO	$474,908	$1,461,655	$986,747
CT	$3,297,198	$23,096,620	$19,799,422
DE	$2,392,564	$2,402,618	$10,054
FL	$15,305,920	$42,414,640	$27,108,720
GA	$6,015,034	$15,872,970	$9,857,936
IA	$328,829	$464,338	$135,509
IL	$1,363,332	$2,665,297	$1,301,965
IN	$3,950,140	$4,878,076	$927,936
KS	$867,276	$1,144,596	$277,320
KY	$18,752,430	$24,658,420	$5,905,990
LA	$959,352,500	$969,094,200	$9,741,700
MA	$2,647,111	$4,378,599	$1,731,488
MD	$7,123,649	$13,698,820	$6,575,171
MI	$2,662,786	$9,822,356	$7,159,570
MN	$5,610	$37,407	$31,798
MO	$2,731,713	$4,522,568	$1,790,855
MS	$22,292,780	$51,854,500	$29,561,720
MT	$1,947	$71,168	$69,221
NC	$12,292,140	$27,096,980	$14,804,840
NE	$234,781	$348,338	$113,557
NH	$839,257	$364,551	-$474,707
NJ	$273,714,600	$247,391,900	-$26,322,700
NM	$390,758	$1,134,124	$743,367
NV	$36,652	$325,155	$288,502
NY	$111,067,600	$108,304,100	-$2,763,500
OH	$2,819,269	$3,274,491	$455,222

Table 5.6 continued

OK	$3,044,045	$3,626,120	$582,075
OR	$235,359	$732,830	$497,471
PA	$79,663,810	$67,506,410	-$12,157,400
RI	$1,239,834	$1,705,389	$465,555
SC	$5,109,030	$13,677,960	$8,568,930
TN	$55,065,840	$50,429,740	-$4,636,100
TX	$43,268,330	$83,154,220	$39,885,890
UT	$91,195	$1,214,148	$1,122,953
VA	$1,566,461	$12,383,320	$10,816,859
VT	$270,266	$341,405	$71,140
WA	$11,222,840	$11,475,810	$252,970
WI	$255,049	$1,274,774	$1,019,725
WV	$4,797,486	$5,150,959	$353,473

5.5 Conclusion

In the last two decades a mixed methodology of statistical methods, data analytics, and machine learning techniques have become more prominent in flood loss estimation research. Researchers have been using historical data to develop various types of empirical models to derive insights from those data. Following a similar approach, this paper has developed a regression model that can predict the county level annual NFIP payout with reasonable accuracy. The model was developed using historical flood insurance claims data between 2016 and 2020, and the performance of the model was tested on the 2021 data. The model predicted the annual NFIP claim originating in a county in 2021 with a mean absolute error of $831,178. It has been shown that the model was able to predict the flood insurance claims that came out of Louisiana after it was devastated by hurricane Ida in 2021 with reasonable accuracy.

The proposed model provides a cheaper alternative for estimating the insured flood losses in the U.S. Once the insured flood losses are estimated, it can be used to predict the extent of uninsured losses in a county. Therefore, the proposed model can be useful to disaster management agencies for estimating the cost of future floods to the primary insurer and the government. It should be noted that recursive modeling is required to keep the model grounded on the latest data.

As more data becomes available, it should be incorporated into the model to fine tune the model's outcomes.

There are certain limitations to this research. First, linear regression requires normality assumption for the residuals. In this case, the response variable, i.e., the annual NFIP payout did not follow a normal distribution. This is a violation of the normality assumption. However, since the objective of this paper is not to make causal inference based on regression coefficients, the normality assumption was neglected as the developed model produced sufficient prediction accuracy on the training and test set. Second, the flood exposure data used in the model is static in nature. The data was only available for the year 2022. In absence of better alternatives, it was assumed that the flood exposure remained constant between 2016 and 2022. This might not be true given the adverse effects of climate change. Lastly, the model requires accurate estimation of flood damage as it is used as a predictor. Flood damage calculation is a challenging task. These limitations provide the ground for future research. In future, non-parametric machine learning models can be developed so that they do not require the normality assumption of the response variable. Also, the data driven models could be combined with existing hydrologic models to create more robust results.

6. CONCLUSION

This chapter summarizes the research results, contributions to the body of knowledge, research limitations, and recommendations for future research.

6.1 Summary and Contributions to the Body of Knowledge

The National Flood Insurance Program (NFIP) provides affordable flood insurance to more than four million households in the U.S. The NFIP faces significant financial challenges, being heavily indebted to the U.S. treasury (Grigg 2019) and anticipating further debt increase due to the escalating frequency and severity of natural hazards (CBO 2017). One key factor contributing to the insolvency issue is the low flood insurance take-up rate in the U.S. (Kousky 2011, Michel-Kerjan et al. 2012, Kousky et al. 2018), with FEMA's designated 100-year flood zones largely underestimating actual flood risk (1st Street Foundation 2020). Moreover, the NFIP premiums are not risk-based, and bureaucratic limitations hinder the program's ability to adjust premiums adequately. Additionally, the NFIP serves as the insurer of last resort, sometimes covering households deemed uninsurable by private flood insurers, which creates further pressure on the program. The presence of asymmetrically used information between the insurer and insured adds to the challenges, with a substantial number of claims originating from repetitive loss properties (Grigg 2019). To address all these challenges and to keep the NFIP solvent, this research has answered three fundamental questions that policy makers need to answer. They are (1) how to increase the NFIP take-up rate, (2) how to reduce the likelihood of large payouts in the future, and (3) how to predict the future payouts. Answering these three questions can help the policy makers to take additional steps, design new floodplain policies, etc., that can keep the NFIP running in the long term.

It has been claimed previously that the availability of post-disaster federal assistance reduces the motivation of the households to insure themselves against future floods. This event is popularly known as Charity Hazard. Researchers have found conflicting evidence of the presence of charity hazard in the U.S. flood insurance market. To address that question, this research has utilized a propensity score-based approach while utilizing relatively new data to compare the post-disaster flood insurance enrollment. The used approach is beneficial as it compares the outcome variables while balancing the effects of the underlying confounding factors such as flood damage, new federal mortgages, etc. It has been found that the receivers of the financial assistance from the government are more inclined to purchase flood insurance than those that did not receive any financial assistance. The research outcomes showed that the average annual difference in the flood insurance enrollment between counties that received IHP assistance and their counterparts that did not receive IHP assistance was 5.2% for the number of NFIP policies and 4.6% for the total insured value of those NFIP policies. Additionally, this research has developed a dose-response function that shows the change in the flood insurance enrollment based on different levels of IHP assistance. The developed dose-response function found that for every 1000 households in a county that received IHP payout, the percentage increase in the number of NFIP policies was 3.41%. On the other hand, for each million-dollar IHP payout in a county, the total insured value of the NFIP policies increased by 1.96% in the following year. Therefore, IHP assistance did not crowd out the demand for flood insurance. In other words, Charity Hazard does not exist in the U.S. flood insurance market.

This knowledge is significant as it can change the way federal assistance is thought of. Currently, IHP is considered a program that helps disaster survivors. The rising cost of IHP and its negative externalities such as crowding out the demand for flood insurance, etc., have been widely discussed. However, this research proved that IHP can help in improving the NFIP take-up rate, which is urgently required to keep the NFIP solvent. The evidence can help the policy makers to design post-disaster financial assistance more effectively so that it reaches the most at-risk section of the community. Previous researchers have found that disaster assistance can be politically motivated (Kousky et al. 2018) and contains significant procedural barriers that can cause disadvantages for the vulnerable population (Hooks and Miller 2006, Grube et al. 2018). Therefore, there is a need to make the IHP assistance more inclusive since it has a positive effect on the flood insurance enrollment. If the IHP assistance is designed to reach the vulnerable population, based on the research results, it will increase their participation in the NFIP, which will increase the NFIP take-up rate and revenue. At the same time, having flood insurance will make the vulnerable communities more resilient to future flood events.

Moreover, the positive effect of IHP assistance on the NFIP enrollment can go a long way. If through IHP assistance the NFIP enrollment can be increased, it will help the government in controlling uninsured flood losses as more people will have flood insurance. Due to climate change, the frequency and severity of natural hazards are expected to increase. Therefore, without the increase in NFIP enrollment, the uninsured losses will increase as more people will get affected by natural hazards. Hence, if IHP assistance can help in improving the NFIP enrollment, the feedback effect (i.e., the effect of higher NFIP take-up rate on IHP assistance) will control the cost of running IHP in the future. The outcomes of this research can facilitate that analysis since it has quantified the extent to which IHP assistance influenced NFIP enrollment.

In the next part, this research has developed a causal model that has quantified the relationship between the flood risk factors and the flood insurance payout. It has been explained before that in order to keep the NFIP solvent, it is essential to reduce the likelihood of large payouts in the future. This research argued that it can be achieved through flood risk reduction. But prior to undertaking the risk reduction initiatives, it is necessary to understand how flood risk factors affect the flood insurance payouts. This research has derived those relationships based on historical data. The Mixed Effects Regression model demonstrates how different flood risk factors influence the flood insurance payout in a county. For instance, it has been found that increasing the flood exposure by $21.7 billion in a county will increase the average annual NFIP payout by $104.8 million. If the infrastructure vulnerability, measured in terms of per capita public assistance payout, increases by $184, the average annual NFIP payout will increase by $84.3 million. Social vulnerability decreases people's ability to attenuate the risk of natural hazards. Therefore, higher social vulnerability potentially leads to higher flood damage and subsequently higher flood insurance payout. If social vulnerability in a county, as quantified by the CDC, increases by 0.27, the average annual NFIP payout will increase by $77 million. Lastly, mobile homes are more prone to damage from natural hazards. If the number of mobile homes in a county increases by 7704, the average annual NFIP payout will increase by $90.4 million. These relationships are one of the key contributions of this research to the body of knowledge as it bridges the gap between flood risk factors and flood insurance payout.

These relationships can help in forecasting future NFIP payout based on different counterfactuals. For instance, according to FEMA's National Risk Index, New Orleans – the largest city of Louisiana currently has a combined flood exposure of $74.7 billion ($45.5 billion from coastal flood and $29.2 billion from river flood) (Zuzak et al. 2021). The 1st Street Foundation predicted that in 30 years the number of properties within the 100-year flood zone in New Orleans will increase by approximately 67% from 66131 to 110236 (1st Street Foundation 2023). Assuming that the coastal flood exposure will increase at the same rate, the combined flood exposure in New Orleans, Louisiana will become \$124.5 billion by 2030, i.e., an increase by \$49.8 billion. Based on the developed causal model, this increase in flood exposure will increase the average annual NFIP payout by \$241 million ($\frac{\$49.8\ billion}{\$21.7\ billion} \times \$104.8\ million$) if everything else remains the same as present. It should be noted that this is a conservative estimate since it does not take into account the price inflation of homes in the calculation. The S&P CoreLogic Case-Shiller U.S. National Home Price NSA Index monitors the fluctuations in the value of the U.S. residential housing market by monitoring single-family home purchase prices. The S&P Dow Jones Indices LLC (2023) records show the index has increased by approximately 4 times from 76.4 in January 1993 to 305.1 in May 2023. If the increase in the housing price follows the same historical trend, it will become 4 times of the present price in the next 30 years. If the housing price increases by 4 times, the combined effect of housing price inflation and increased flood exposure by 67% will increase the expected annual NFIP payout by \$2.04 billion. The research results facilitate this type of analysis.

On the other hand, the relationships can be utilized to quantify the impact of a risk reduction strategy or policy on the annual NFIP payout. Before implementing a strategy, it is essential to estimate the benefit and cost associated with that strategy. Benefits from a strategy can be calculated as the reduction of possible losses due to the implementation of the strategy (Bhattacharyya et al. 2021). The causal model developed in this research can be useful for that benefit analysis. For instance, if property buyout reduces the exposure by 1% in New Orleans, it will reduce the expected annual NFIP payout by \$20 million, i.e., 1% of \$2.04 billion. Additionally, if multiple policies are planned to mitigate different flood risk factors, the causal model can be utilized to find the optimal mix of different policies and strategies that can maximize the benefits in terms of reducing future payout under different constraint such as budget, etc.

Lastly, this research has developed a predictive model that can predict the future annual NFIP payout in a county with reasonable accuracy. This model was utilized to predict the NFIP claims after Hurricane Ida and as it has been explained previously, it predicted the claim amount with very high accuracy. The predictive model can be used by FEMA and other public agencies to understand the extent of NFIP claims in future year, which can help them to improve their financial preparedness for future disasters. Moreover, by estimating the flood insurance claim amount, the developed predictive model can also help in estimating the uninsured loss amount.

To summarize, this research was intended to make and/or keep the NFIP financially viable in the long term so that the U.S. federal government can continue to provide subsidized flood insurance to homeowners. To do that, this research identified three fundamental questions that required answering. They are (1) how to improve the NFIP take-up rate so that more revenue can be generated, (2) how to reduce the likelihood of large payouts so that the program remains solvent despite the increasing frequency of natural hazards, and (3) how to predict future payout so that NFIP can be financially prepared for that. The research outcomes showed that contrary to the popular belief, post-disaster federal assistance can increase the NFIP enrollment. This insight can have a significant impact in the way post-disaster federal assistance is planned and disbursed. If it reaches the most at-risk population and helps them to get back on their feet after a disaster, the results indicate that they will be encouraged to buy and maintain flood insurance in the future. There are two benefits from this. First, having flood insurance increases the ability to recover from disasters. Therefore, if the most at-risk population are insured by the NFIP, it will increase their resilience for future flood events. Secondly, as more people purchase flood insurance, it will generate more revenue, which will help the NFIP to remain financially viable and solvent. At the same time, the NFIP needs to plan and undertake different risk reduction strategies, policies, or a combination of the two. The research results will help in planning those strategies. This research contributes to the body of knowledge by quantifying the causal relationships between the identified flood risk factors and flood insurance payout, which has not been done previously. Lastly, the predictive model can help the NFIP to prepare for future flood insurance claims.

Based on the analysis, this research recommends a number of policies that can be implemented to keep the NFIP running in the long term. They are as follows.

(1) The NFIP premiums should be risk-based. As noted by the Department of Homeland Securities, the majority of the flood maps currently used by the FEMA are inadequate in quantifying the flood risk accurately (DHS 2017). Therefore, a Catastrophe Risk Model-based actuarial approach is required that will calculate the flood risk for each property individually and adjust the premium accordingly. When analyzing the flood risk, future hazard scenarios based on various climate change situations should also be considered. Moreover, the premiums should be revised time to time so that it reflects the evolving flood risk of a property.

(2) Since making the flood insurance premiums risk-based can increase the cost of the premium, the NFIP should subsidize the socially vulnerable population in addition to the existing

subsidies. Having a flood insurance can reduce the social vulnerability of a community as it augments the community's ability to recover from a disaster and this research has shown that reduced vulnerability reduces the expected flood insurance payout. Therefore, the benefit of the subsidy can be realized in terms of reduction of future payout.

(3) Risk reduction is necessary. The NFIP should spend a part of the proceedings to finance risk reduction and resilience. Since this research recommends that the NFIP premiums be risk-based, by reducing the risk, the NFIP can charge lesser premium to homeowners. The reduced premium can boost the demand for flood insurance, which will subsequently help in improving the flood insurance take-up rate. Through risk reduction, the NFIP can reduce the likelihood of future payout and increase the demand for flood insurance. This will keep the NFIP solvent in the long term. Risk reduction initiatives such as dams and levees along with risk-based premiums can improve the private participation in the U.S. flood insurance market. As noted by Kousky (2018), private insurers struggle to keep up with the subsidized rate that NFIP charges. If NFIP premiums are risk-based, they should be comparable with premiums charged by private insurers. This can help to improve private participation and therefore, improve risk sharing between public and private entities.

(4) Disaster impacts are disproportionate. The losses suffered by racial and ethnic minority groups, poor communities, etc., are generally higher due to lack of inherent resilience. As a result, disasters often increase social vulnerability. Since it has been found that post-disaster federal assistance encourages the disaster survivors to insure themselves against future floods, assistance could be designed so that it reaches the socially vulnerable population. It has been explained before that encouraging the socially vulnerable population to insure themselves can have long term benefits as it reduces the likelihood of large payouts by reducing social vulnerability.

(5) The existing debts of the NFIP to the U.S. treasury should be written off. The NFIP pays hundreds of millions each year as interest on its debts. This money can be spent on funding risk reduction and resilience initiatives. The risk reduction and resilience can reduce large future payouts thus keeping the program solvent. The money could also be spent on transferring additional flood risk through further purchasing reinsurance and CAT bonds.

(6) The government should explore the possibility of mandating flood insurance for all properties located in 100-year floodplains, regardless of their mortgage situation. Currently, flood

insurance is only required for properties in a 100-year flood zone with active federal mortgages, leading to homeowners discontinuing insurance once their mortgages are paid off. 1st Street Foundation estimated that in 2020, there were 14.6 million properties within 100-year flood zones and an additional 7.2 million between 100-year and 500-year flood zones (1st Street Foundation 2020). By mandating insurance for properties within 100-year flood zones and extending the existing regulation to 500-year flood zones, new construction in floodplains could be discouraged, and the number of NFIP policies could be increased significantly, helping to keep the program solvent in the long term.

(7) Lastly, awareness programs are necessary to improve new homeowners' understanding about the present and future flood risk in a location. As explained previously, new constructions in floodplains in many U.S. cities are increasing and many homeowners are not fully informed about the flood risk at those locations (1st Street Foundation 2020). Therefore, there is a need to increase flood awareness among the new homeowners so that they avoid buying homes that are at risk of flooding. The federal government can also enforce certain regulations on real estate companies that force them to disclose previous flood events, current, and future flood risk at the location so that homeowners are fully aware of the existing and future flood risk at their homes.

6.2 Research Limitations

Like any research, there are certain limitations to this research. They are listed below.

(1) First of all, this research has produced three macro-level models that were developed at the county level. Like any macro-level analysis, this research does not take into account the micro-level nuances such as household income, education, etc. Instead, this research has considered those variables at macro-level. Evidently, some information was lost due to this aggregation. However, micro-level models are expensive and require vast data collection, which is the reason this research chose macro-level analysis.

(2) Like any data-driven analysis, the outcomes of this research are dependent on the data used in conducting the analysis. As more data becomes available, they should be incorporated into the models and that might change some of the research outcomes.

(3) The research has concluded that post-disaster federal assistance encourages NFIP enrollment. A part of that increase in enrollment is caused by the federal requirement of maintaining flood

insurance after receiving federal assistance to keep the recipients eligible for future assistance. However, the current research did not distinguish the effect of regulatory purchase from voluntary purchase. Therefore, how much NFIP enrollment was voluntary is unknown from this current research outcomes. To distinguish that effect, micro-level analysis is required. The collected data did not support that analysis.

(4) This research has considered nine confounding variable. Ideally, the propensity score should balance all the confounding variables. Out of the nine confounding variables, the derived propensity score achieved balance for eight of them that excluded the IHP approval rate. This can be considered another limitation of the analysis. Additionally, the list of confounding variables is not exhaustive. Other confounding variables can also be considered.

(5) This research requires accurate estimation of flood damage. Despite several available tools and methodologies, accurate flood damage estimation is a challenging task, which is a limitation to this research. It should be noted that this research did not calculate the damage from previous flood events. It rather used NOAA's flood damage estimation in the analysis. Therefore, the any inaccuracy in NOAA's flood damage estimation might have affected the outcomes of this research.

(6) The flood exposure data used in this research is static in nature. The data was collected from FEMA's National Risk Index and was only available for one year. In absence of better alternatives, this research has assumed that the flood exposure remained constant between 2016 and 2021, which might not be the actual situation given the rapid changes in flood risk due to climate change.

(7) The list of flood risk factors is not exhaustive. As explained previously, this research focused on the flood risk factors (flood exposure, infrastructure vulnerability, social vulnerability, community resilience, and the number of mobile homes) that can be controlled through human intervention. There are other factors that might influence the flood risk in a county such as soil conditions, terrain, elevation, etc. Since these factors cannot be controlled through human interventions, they were kept out of scope of this research.

(8) The derived causal relationships are not permanent. Due to the dynamic nature of underlying control variables, the empirical relationships between flood risk factors and flood insurance payout might change in future. Therefore, the models should be updated periodically as more data becomes available.

(9) It has been explained previously that the derived causal estimates can be used to plan for structural measures such as building new flood protection infrastructure, etc., to reduce flood insurance claims. However, structural measures such as dams and levees sometimes create an illusion of false safety among the populations protected by those dams and levees, which leads to risk compensation behavior (Kundzewicz et al. 2018). Risk compensation is a popular notion in psychology and economics that claims people adjust their behavior based on the perceived risk. These structural measures often encourage developers to build new homes behind these levees. Again, new construction in the floodplain further increases the overall flood exposure, which increases flood insurance payout. The statistical models developed in this research does not take into account this risk compensation behavior of households that leads to a feedback loop between flood protection and flood exposure.

(10) The derived prediction model required NFIP claim count estimation as it is one of the predictors for the regression models. This research has taken advantage of a model that has been proposed by Ghaedi et al. (2022). Their model could predict the number of annual flood insurance claims in a county with reasonable accuracy. Hence, it has been assumed that the number of NFIP claims can be estimated and used for the prediction purpose.

(11) The prediction models were trained to predict annual NFIP payout up to \$500 million. Therefore, it is not recommended to use predictions that are higher than \$500 million.

6.3 Recommendations for Future Research

The limitations of this research can provide directions for future research on this topic. Since macro-models lack micro-level nuances, in future similar analysis could be conducted using household level data to see if the research outcomes differ due to the change in approach. The analysis of charity hazard can be conducted using propensity score-based methods for individual households. This has not been done in the past and hence, it can be explored. Researchers can also use choice experiments to investigate, which factors influence the demand for flood insurance at the micro level. Similarly, the causal model and the predictive model can be developed for each household. However, it will be challenging to collect the household level data to develop those models.

In future, research can be conducted to see how post-disaster assistance can be designed to alleviate social vulnerability. As reflected in previous literature, there are certain procedural barriers that hinder the access of these assistance to the vulnerable populations. It is worth exploring how those barriers can be broken so that the at-risk population receives the benefits of the assistance program. The benefits from this can go a long way as reduced social vulnerability will further reduce the extent of flood insurance payouts.

Lastly, flood risk reduction is a multistakeholder process. Regarding flood resilience it has been established that an individual organization or stakeholder cannot successfully understand and resolve flood risk (Australian Public Service Commission 2012). For instance, if the government decides to build flood protection infrastructure, that can lead to new construction in the floodplain due to risk compensation behavior from the developers and households. This drastically increases the number of properties exposed to flood hazards. On the other hand, without any structural measures, the true flood risk based on actuarial estimations at a location will increase with time, which might make the flood insurance premiums unaffordable in the future. If the premiums are unaffordable, it will certainly reduce the demand for flood insurance. Therefore, solving this challenge will require a collaborative effort among all the stakeholders. To facilitate that, a multistakeholder collaboration framework that brings all the stakeholders (households, insurers, reinsurers, CAT bond agencies, and government) on board in planning and financing flood risk reduction and resilience initiatives should be explored. It will be interesting to explore if a win-win situation can be created for all the stakeholders involved in the collaboration. Additionally, models are required to understand the impact of policies such as raising the cost of premium for new constructions in floodplain, revising the zoning code to convert parts of the floodplain into greenspace, etc.

REFERENCES

1. 1st Street Foundation. (2020). The First National Flood Risk Assessment: Defining America's Growing Risk I

2. 1st Street Foundation. (2023). Risk Factor. https://riskfactor.com/city/new-orleans-louisiana/2255000_fsid/flood#current_&_future_risk accessed July 28, 2023

3. Aarts, E., Aarts, E. H., & Lenstra, J. K. (Eds.). (2003). *Local search in combinatorial optimization*. Princeton University Press.

4. Abotaleb, I. S., & El-adaway, I. H. (2018). First attempt toward a holistic understanding of the interdependent rippled impacts associated with out-of-sequence work in construction projects: System dynamics modeling approach. *Journal of Construction Engineering and Management*, *144*(9), 04018084.

5. Akabas S., (2014). Possible Reforms to the National Flood Insurance Program. Available from: https://bipartisanpolicy.org/blog/possible-reforms-national-flood-insurance-program/

6. Alireza, V., Mohammadreza, Y., Zin, R. M., Yahaya, N., & Noor, N. M. (2014). An enhanced multi-objective optimization approach for risk allocation in public–private partnership projects: a case study of Malaysia. *Canadian Journal of Civil Engineering*, *41*(2), 164-177.

7. Allen, M., Gillespie-Marthaler, L., Abkowitz, M., & Camp, J. (2020). Evaluating flood resilience in rural communities: a case-based assessment of Dyer County, Tennessee. *Natural Hazards*, *101*(1), 173-194.

8. Aman, D. D., and Yarnal, B. (2010). "Home sweet mobile home? Benefits and challenges of mobile home ownership in rural Pennsylvania." Appl. Geogr., 30(1), 84–95.

9. Andor, M. A., Osberghaus, D., & Simora, M. (2020). Natural disasters and governmental aid: Is there a charity hazard?. Ecological Economics, 169, 106534.

10. Armal, S., Porter, J. R., Lingle, B., Chu, Z., Marston, M. L., & Wing, O. E. (2020). Assessing property level economic impacts of climate in the US, new insights and evidence from a comprehensive flood risk assessment tool. Climate, 8(10), 116.

11. ARTEMIS (2021) (https://www.artemis.bm/deal-directory/?_sft_perils=u-s-flood) accessed March 14, 2021

12. ASCE. (2021). 2021 Report Card for America's infrastructure. American Society of Civil Engineers.

13. Ashley, R., Walker, L., D'Arcy, B., Wilson, S., Illman, S., Shaffer, P., ... & Chatfield, P. (2015, August). UK sustainable drainage systems: past, present and future. In Proceedings of the Institution of Civil Engineers-Civil Engineering (Vol. 168, No. 3, pp. 125-130). Thomas Telford Ltd. 10.1680/cien.15.00011

14. Atreya, A., Hanger, S., Kunreuther, H., Linnerooth-Bayer, J., & Michel-Kerjan, E. (2015). A comparison of residential flood insurance markets in 25 countries. Wharton Risk Management and Decision Processes Center, University of Pennsylvania, Philadelphia, June, 28.

15. Atreya, A., Ferreira, S., & Kriesel, W. (2013). Forgetting the flood? An analysis of the flood risk discount over time. *Land Economics*, *89*(4), 577-596.

16. Austin, P. C. (2011a). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical statistics*, *10*(2), 150-161.

17. Austin P. C. (2011b). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate behavioral research*, *46*(3), 399–424. https://doi.org/10.1080/00273171.2011.568786

18. Australian Public Service Commission. (2012). Tackling wicked problems: A public policy perspective

19. Baker, D., Hamshaw, K., and Beach, C. (2011). "A window into park life: Findings from a resident survey of nine mobile home park communities in Vermont." J. Rural Community Dev., 6(2), 53–70.

20. Baker, D., Hamshaw, S. D., & Hamshaw, K. A. (2014). Rapid flood exposure assessment of Vermont mobile home parks following Tropical Storm Irene. Natural Hazards Review, 15(1), 27-37.

21. Batica, J., & Gourbesville, P. (2016). Resilience in flood risk management–a new communication tool. Procedia Engineering, 154, 811-817. doi: 10.1016/j.proeng.2016.07.411

22. Bauer, H., Nowak, D., & Herbig, B. (2018). Aging and Cardiometabolic Risk in European HEMS Pilots: An Assessment of Occupational Old-Age Limits as a Regulatory Risk Management Strategy. Risk Analysis, 38(7), 1332-1347.

23. Beyer, H. G. (2001). *The theory of evolution strategies*. Springer Science & Business Media.

24. Bhattacharyya, A., Morshedi, M., & Hastak, M. (2023). A clustering-classification approach in categorizing vulnerability of roads and bridges using public assistance big data. International Journal of Disaster Risk Reduction, 84, 103448.

25. Bhattacharyya, A., Yoon, S., & Hastak, M. (2021). Optimal Strategy Selection Framework for Minimizing the Economic Impacts of Severe Weather Induced Power Outages. International Journal of Disaster Risk Reduction, 102265. https://doi.org/10.1016/j.ijdrr.2021.102265

26. Bhattacharyya, A., Yoon, S., and Hastak, M. (2021). Economic Impact Assessment of Severe Weather–Induced Power Outages in the US. *Journal of Infrastructure Systems*, 27(4), 04021038. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000648

27. Bizikova, L., Robinson, J., & Cohen, S. (2007). Linking climate change and sustainable development at the local level.

28. Boamah, S. A., Armah, F. A., Kuuire, V. Z., Ajibade, I., Luginaah, I., & McBean, G. (2015). Does previous experience of floods stimulate the adoption of coping strategies? Evidence from cross sectional surveys in Nigeria and Tanzania. Environments, 2(4), 565-585.

29. Boreczky, R. (2006), "US program maps out flood coverage", Washington Times, May 12, p. F01.

30. Botzen, W. W., & van den Bergh, J. C. (2012). Risk attitudes to low-probability climate change risks: WTP for flood insurance. *Journal of Economic Behavior & Organization*, *82*(1), 151-166.

31. Botzen, W. J., Aerts, J. C. J. H., & van den Bergh, J. C. (2009). Dependence of flood risk perceptions on socioeconomic and objective risk factors. Water resources research, 45(10).

32. Brabhaharan, P. (2021). Integrated Wellington region land transport resilience study. Bulletin of the New Zealand Society for Earthquake Engineering, 54(2), 163-175. https://doi.org/10.5459/bnzsee.54.2.163-175

33. Bradt, J. T., Kousky, C., and Wing, O. E. (2021). Voluntary purchases and adverse selection in the market for flood insurance. Journal of Environmental Economics and Management, 110, 102515.

34. Breiman, L. (2001). Statistical modeling: the two cultures. Stat. Sci. 16, 199–231. doi:10.1214/ss/1009213726

35. Browne, M. J., and Hoyt, R. E. (2000). The Demand for Flood Insurance: Empirical Evidence. Journal of Risk and Uncertainty, 20(3), 291–306. http://www.jstor.org/stable/41760986

36. Campbell, N., Roper-Fetter, K., & Yoder, M. (2020). Principles of risk communication: A guide to communicating with socially vulnerable populations across the disaster lifecycle. *Natural Hazards Center, University of Colorado: Boulder, CO, USA*.

37. Carter, N. T., Horn, D. P., Brown, J. T., Boyd, E., Lipiec, E., Stubbs, M., & Normand, A. E. (2019). Flood resilience and risk reduction: Federal assistance and programs. Report R45017. Washington, DC: Congressional Research Service.

38. CBO (2017) The National Flood Insurance Program: Financial Soundness and Affordability https://www.cbo.gov/system/files/115th-congress-2017-2018/reports/53028-nfipreport2.pdf accessed June 17, 2021

39. CCRIF (2021a) (http://www.ccrif.org/) accessed March 14, 2021.

40. CCRIF (2021b) (https://www.ccrif.org/aboutus/ccrif-spc-payouts) accessed March 14, 2021

41. Chappell, W.F., Forgette, R.G., Swanson, D.A. and van Boening, M.V. (2007), “Determinants of government aid to Katrina survivors: evidence from survey data”, Southern Economic Journal, Vol. 74 No. 2, pp. 344-62.

42. Choi, J., Deshmukh, A., and Hastak, M. (2019). Seven-layer classification of infrastructure to improve community resilience to disasters. Journal of infrastructure systems, 25(2), 04019012.

43. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Academic press.

44. Colgan, C. S. (2017). Financing natural infrastructure for coastal flood damage reduction. Lloyd’s Tercentenary Research Foundation, London.

45. Collischon, M., & Eberl, A. (2020). Let’s talk about fixed effects: Let’s talk about all the good things and the bad things. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *72*(2), 289-299.

46. Congressional Research Service (2021) A Brief Overview of FEMA’s Public Assistance Program

47. CRS (2022). Flood Buyouts: Federal Funding for Property Acquisition. https://crsreports.congress.gov/product/pdf/IN/IN11911 accessed July 8, 2023

48. Cutter S, Boruff B, Shirley W (2003) Social vulnerability to environmental hazards. Soc Sci Q 84(1):242–261. doi:10.1111/1540-6237.8402002

49. Cutter, S. L., Ash, K. D., and Emrich, C. T. (2014). The geographies of community disaster resilience. Global environmental change, 29, 65-77.

50. Dandapat, K., and Panda, G. K. (2017). Flood vulnerability analysis and risk assessment using analytical hierarchy process. Modeling Earth Systems and Environment, 3(4), 1627-1646.

51. Davenport, F. V., Burke, M., & Diffenbaugh, N. S. (2021). Contribution of historical precipitation change to US flood damages. *Proceedings of the National Academy of Sciences*, *118*(4). https://doi.org/10.1073/pnas.2017524118

52. Davlasheridze, M., & Miao, Q. (2019). Does governmental assistance affect private decisions to insure? An empirical analysis of flood insurance purchases. Land Economics, 95(1), 124-145.

53. Deopersad, C., Persaud, C., Chakalall, Y., Bello, O., Masson, M., Perroni, A., ... & Invest, B. I. D. (2020). Assessment of the Effects and Impacts of Hurricane Dorian in the Bahamas. *Inter-American Development Bank*, 219.

54. Desai, S., and Ouarda, T. B. (2021). Regional hydrological frequency analysis at ungauged sites with random forest regression. Journal of Hydrology, 594, 125861.

55. DHS (2017). FEMA Needs to Improve Management of its Flood Mapping Programs. Office of Inspector General. https://s3.documentcloud.org/documents/4066233/OIG-17-110-Sep17.pdf accessed May 17, 2023

56. Diakakis M (2014) An inventory of flood events in Athens, Greece, during the last 130 years. Seasonality and spatial distribution. J Flood Risk Manag 7:332–343.

57. Dieperink, C., Mees, H., Priest, S., Ek, K., Bruzzone, S., Larrue, C., & Matczak, P. (2018). Managing urban flood resilience as a multilevel governance challenge: An analysis of required multilevel coordination mechanisms. *Ecology and Society, 23*(1). doi:10.2307/26799053

58. Dixon, L. S., Clancy, N. B., Bender, B., Kofner, A., Manheim, D., & Zakaras, L. (2013). *Flood Insurance in New York City Following Hurricane Sandy*. RAND.

59. Dixon, L. S. (2007). *The lender-placed flood insurance market for residential properties* (Vol. 468). Rand Corporation.

60. Drakes, O., Tate, E., Rainey, J., & Brody, S. (2021). Social vulnerability and short-term disaster assistance in the United States. International Journal of Disaster Risk Reduction, 53, 102010.

61. Dubbelboer, J., Nikolic, I., Jenkins, K., & Hall, J. (2017). An agent-based model of flood risk and insurance. Journal of Artificial Societies and Social Simulation, 20(1). http://jasss.soc.surrey.ac.uk/20/1/6.html

62. Efendić, E. (2021). How do people judge risk? Availability may upstage affect in the construction of risk judgments. Risk analysis, 41(11), 2003-2015.

63. Emmert-Streib, F., & Dehmer, M. (2021). Data-driven computational social network science: predictive and inferential models for web-enabled scientific discoveries. Frontiers in big Data, 4, 591749.

64. Emrich, C. T., Tate, E., Larson, S. E., and Zhou, Y. (2020). Measuring social equity in flood recovery funding. Environmental Hazards, 19(3), 228-250.

65. Ezell, B. C. (2007). Infrastructure vulnerability assessment model (I-VAM). Risk Analysis: An International Journal, 27(3), 571-583.

66. FEMA (2021) National Risk Index Technical Documentation https://www.fema.gov/sites/default/files/documents/fema_national-risk-index_technical-documentation.pdf accessed April 1, 2022

67. FEMA (2022) Risk Rating 2.0: Equity in Action. https://www.fema.gov/flood-insurance/risk-rating accessed March 21, 2022

68. FEMA, "Notice of Maximum Amount of Assistance Under the Individuals and Households Program," 84 Federal Register 55324, October 16, 2019, https://www.govinfo.gov/content/pkg/FR-2019-10-16/pdf/2019-22471.pdf

69. FEMA, National Flood Insurance Program Report to Congress on Reinsuring NFIP Insurance Risk and Options for Privatizing the NFIP, August 13, 2015, p. 85.

70. FEMA. (2021). Flood Insurance Manual. https://www.fema.gov/sites/default/files/documents/fema_nfip-all-flood-insurance-manual-apr-2021.pdf accessed July 9, 2023

71. Feng, K., Wang, S., Li, N., Wu, C., & Xiong, W. (2018). Balancing public and private interests through optimization of concession agreement design for user-pay PPP projects. Journal of Civil Engineering and Management, 24(2), 116-129. https://doi.org/10.3846/jcem.2018.455

72. Flavelle, C. (July 31, 2019). Homes Are Being Built the Fastest in Many Flood-Prone Areas, Study Finds. The New York Times. https://www.nytimes.com/2019/07/31/climate/climate-change-new-homes-flooding.html accessed on June 16, 2023

73. Fontaine, S., Yang, Y., Qian, W., Gu, Y., & Fan, B. (2020). A unified approach to sparse Tweedie modeling of multisource insurance claim data. Technometrics, 62(3), 339-356.

74. Frazier, T., Boyden, E. E., and Wood, E. (2020). Socioeconomic implications of national flood insurance policy reform and flood insurance rate map revisions. Natural Hazards, 103(1), 329-346.

75. Gallagher, J. (2014). Learning about an infrequent event: Evidence from flood insurance take-up in the United States. *American Economic Journal: Applied Economics*, 206-233.

76. Garrett, T. A., & Sobel, R. S. (2003). The political economy of FEMA disaster payments. Economic inquiry, 41(3), 496-509.

77. Geldof, T., Popovic, D., Van Damme, N., Huys, I., & Van Dyck, W. (2020). Nearest neighbour propensity score matching and bootstrapping for estimating binary patient response in oncology: a Monte Carlo simulation. *Scientific reports*, *10*(1), 964.

78. GFDRR and World Bank Group. (2014) Financial protection against disasters: an operational framework for disaster risk financing and insurance. World Bank.

79. Ghaedi, H., Reilly, A. C., Baroud, H., Perrucci, D. V., and Ferreira, C. M. (2022). Predicting flood damage using the flood peak ratio and Giovanni Flooded Fraction. PloS one, 17(8), e0271230.

80. Ghesquiere, F. and Mahul, O. (2010) Financial protection of the state against natural disasters: A primer. World Bank Policy Research Working Paper Series, Paper 5429.

81. Gianicolo, E. A., Eichler, M., Muensterer, O., Strauch, K., & Blettner, M. (2020). Methods for evaluating causality in observational studies: Part 27 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International*, *117*(7), 101.

82. Glover, F. (1986). Future paths for integer programming and links to artificial intelligence. *Computers & operations research*, *13*(5), 533-549.

83. Grafakos, S., Trigg, K., Landauer, M., Chelleri, L., & Dhakal, S. (2019). Analytical framework to evaluate the level of integration of climate adaptation and mitigation in cities. *Climatic change*, *154*(1), 87-106.

84. Grigg, N. S. (2019). US flood insurance at 50 years: is the public–private partnership working?. Water Policy, 21(3), 468-480.

85. Gurenko, E. and Mahul, O. (2003) Combining insurance, contingent debt and self-retention an optimal corporate risk financing strategy. World Bank Working Paper 3167.

86. Haer, T., Botzen, W. W., Van Roomen, V., Connor, H., Zavala-Hidalgo, J., Eilander, D. M., & Ward, P. J. (2018). Coastal and river flood risk analyses for guiding economically optimal flood adaptation policies: a country-scale study for Mexico. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2121), 20170329.

87. Hallauer, A. M., Tyagi, A., Behrend, G., Bell, D., & Cohen, B. (2019, May). Environmental Impact Bond: An Innovative Financing Mechanism for Enhancing Resilience in the City of Atlanta through Green Infrastructure. In *World Environmental and Water Resources Congress 2019: Emerging and Innovative Technologies and International Perspectives* (pp. 74-84). Reston, VA: American Society of Civil Engineers. https://doi.org/10.1061/9780784482322.008

88. Halter, W., & Mostaghim, S. (2006, July). Bilevel optimization of multi-component chemical systems using particle swarm optimization. In *2006 IEEE International Conference on Evolutionary Computation* (pp. 1240-1247). IEEE.

89. Harvey, C. A., Chacón, M., Donatti, C. I., Garen, E., Hannah, L., Andrade, A., ... & Wollenberg, E. (2014). Climate-smart landscapes: opportunities and challenges for integrating adaptation and mitigation in tropical agriculture. *Conservation Letters*, *7*(2), 77-90.

90. Hegger, D. L., Driessen, P. P., Dieperink, C., Wiering, M., Raadgever, G. T., and van Rijswick, H. F. (2014). Assessing stability and dynamics in flood risk governance. Water Resources Management, 28(12), 4127-4142.

91. Heinz Center (2002). Human Links to Coastal Disasters. Washington D.C.: The H. John Heinz III Center for Science, Economics and the Environment.

92. Hicks RWB, Woods SD (2000) Pollutant load, population growth and land use. Prog: Water Environ Res Found 11:10

93. Hirano, K., & Imbens, G. W. (2004). The propensity score with continuous treatments. Applied Bayesian modeling and causal inference from incomplete-data perspectives, 226164, 73-84.

94. Hochrainer, S. (2006). Catastrophe modeling and simulation. *Macroeconomic Risk Management Against Natural Disasters: Analysis focussed on governments in developing countries*, 105-144.

95. Horn, D. P. (2019) Closing the Flood Insurance Gap (Updated November 26, 2019) https://www.everycrsreport.com/files/20191126_IN10890_fabf08ab36aa046dd22d2172a1ca702900d4c6f7.pdf accessed June 17, 2021

96. Horn, D. P. (2020) National Flood Insurance Program Borrowing Authority (Updated October 2, 2020) https://fas.org/sgp/crs/homesec/IN10784.pdf accessed June 17, 2021

97. Horn. D. P. and Webel. B. (2021). Private Flood Insurance and the National Flood Insurance Program. Congressional Research Service. https://sgp.fas.org/crs/homesec/R45242.pdf accessed November 4, 2022

98. Huber, P. J. (1964). Robust estimation of a location parameter. Annals of Mathematical Statistics 35, 73–101.

99. Huff, A. G., Hodges, J. S., Kennedy, S. P., & Kircher, A. (2015). Evaluation of the food and agriculture sector criticality assessment tool (FASCAT) and the collected data. Risk Analysis, 35(8), 1448-1467.

100. Hwang, S., Park, M., Lee, H. S., Lee, S., & Kim, H. (2015). Post-disaster interdependent built environment recovery efforts and the effects of governmental plans: Case analysis using system dynamics. *Journal of construction engineering and management*, *141*(3), 04014081.

101. Imbens GW. The role of the propensity score in estimating dose–response functions. Biometrika 2000; 87: 706–710.

102. Indaco, A., Ortega, F., & Taşpınar, S. (2019). The effects of flood insurance on housing markets. Cityscape, 21(2), 129-156.

103. Insurance Information Institute "Spotlight on: Flood insurance" (October 6, 2020) (https://www.iii.org/article/spotlight-on-flood-insurance) accessed March 14, 2021.

104. ISDR, U. (2005). Hyogo framework for action 2005-2015: building the resilience of nations and communities to disasters. In *Extract from the final report of the World Conference on Disaster Reduction (A/CONF. 206/6)* (Vol. 380).

105. Jeerangsuwan, T., Said, H., Kandil, A., & Ukkusuri, S. (2012). Optimization application for financial viability evaluation of PPP toll road projects. In *Construction Research Congress 2012: Construction Challenges in a Flat World* (pp. 2329-2338).

106. Jevrejeva, S., Jackson, L. P., Grinsted, A., Lincke, D., & Marzeion, B. (2018). Flood damage costs under the sea level rise with warming of 1.5 C and 2 C. Environmental Research Letters, 13(7), 074014.

107. Jia, L., Wang, Y., & Fan, L. (2014). Multiobjective bilevel optimization for production-distribution planning problems using hybrid genetic algorithm. *Integrated Computer-Aided Engineering*, *21*(1), 77-90.

108. Karl, T., & Koss, W. J. (1984). Regional and national monthly, seasonal, and annual temperature weighted by area, 1895-1983.

109. Kelman, I., and Spence, R. (2004). An overview of flood actions on buildings. Engineering Geology, 73(3-4), 297-309.

110. Knighton, J., Buchanan, B., Guzman, C., Elliott, R., White, E., & Rahm, B. (2020). Predicting flood insurance claims with hydrologic and socioeconomic demographics via machine learning: Exploring the roles of topography, minority populations, and political dissimilarity. Journal of Environmental Management, 272, 111051.

111. Koc, K., and Işık, Z. (2021). Assessment of urban flood risk factors using Monte Carlo analytical hierarchy process. Natural Hazards Review, 22(4), 04021048.

112. Kosuke Imai & David A van Dyk (2004) Causal Inference With General Treatment Regimes, Journal of the American Statistical Association, 99:467, 854-866, DOI: 10.1198/016214504000001187

113. Kousky, C. (2011). Understanding the demand for flood insurance. *Natural Hazards Review*, *12*(2), 96-110.

114. Kousky, C. (2013). Facts about FEMA household disaster aid: examining the 2008 floods and tornadoes in Missouri. Weather, Climate, and Society, 5(4), 332-344.

115. Kousky, C. (2018), Financing Flood Losses: A Discussion of the National Flood Insurance Program. Risk Management and Insurance Review, 21: 11-32. https://doi.org/10.1111/rmir.12090

116. Kousky, C., & Shabman, L. (2012). The realities of federal disaster aid. RFF Issue Brief. Washington, DC: Resources for the Future.

117. Kousky, C. (2011). Understanding the demand for flood insurance. *Natural Hazards Review*, *12*(2), 96-110.

118. Kousky, C., & Michel-Kerjan, E. (2012). Hurricane sandy, storm surge, and the national flood insurance program. *A primer on New York and Jersey.(Issue Brief No. 12–8). Resources for the Future*.

119. Kousky, C., Kunreuther, H., LaCour-Little, M., and Wachter, S. (2020). Flood risk and the US housing market. Journal of Housing Research, 29(sup1), S3-S24.

120. Kousky, C., Kunreuther, H., Lingle, B., & Shabman, L. (2018a). The emerging private residential flood insurance market in the United States. Wharton Risk Management and Decision Processes Center.

121. Kousky, C., Michel-Kerjan, E. O., & Raschky, P. A. (2018b). Does federal disaster assistance crowd out flood insurance?. Journal of Environmental Economics and Management, 87, 150-164.

122. Kriesel, W., and Landry, C. (2004). Participation in the National Flood Insurance Program: An empirical analysis for coastal properties. Journal of Risk and Insurance, 71(3), 405-420.

123. Kundzewicz, Z. W., Hegger, D. L. T., Matczak, P., & Driessen, P. P. J. (2018). Opinion: Flood-risk reduction: Structural measures and diverse strategies. Proceedings of the National Academy of Sciences, 115(49), 12321-12325.

124. Landry, C. E., and Jahan-Parvar, M. R. (2011). Flood insurance coverage in the coastal zone. Journal of Risk and Insurance, 78(2), 361-388.

125. Landry, C. E., Turner, D., & Petrolia, D. (2021). Flood insurance market penetration and expectations of disaster assistance. Environmental and Resource Economics, 79(2), 357-386.

126. Leandro, J., Chen, K. F., Wood, R. R., & Ludwig, R. (2020). A scalable flood-resilience-index for measuring climate change adaptation: Munich city. *Water Research*, *173*, 115502. https://doi.org/10.1016/j.watres.2020.115502

127. Lee, J. (2020). A Tale of Two Counties in the Wake of the 2017 Hurricane Season. International Journal of Disaster Risk Reduction, 101902.

128. Len, N. L. S., Bolong, N., Roslee, R., Tongkul, F., Mirasa, A. K., & Ayog, J. L. (2018). Flood vulnerability of critical infrastructures-review. Malaysian Journal of Geosciences, 2(1), 31-34.

129. Len, N. L. S., Bolong, N., Roslee, R., Tongkul, F., Mirasa, A. K., & Ayog, J. L. (2018). Flood vulnerability of critical infrastructures-review. Malaysian Journal of Geosciences, 2(1), 31-34.

130. Lester, R., & Gurenko, E. (2004). *Rapid onset natural disasters: The role of financing in effective risk management*. The World Bank.

131. Livingston, I. (2021). Ida's impact from the Gulf Coast to Northeast — by the numbers. The Washington Post. https://www.washingtonpost.com/weather/2021/09/03/hurricane-ida-numbers-surge-wind-pressure-damage/ accessed May 16, 2021

132. Locatelli, B., Fedele, G., Fayolle, V., & Baglee, A. (2016). Synergies between adaptation and mitigation in climate change finance. *International Journal of Climate Change Strategies and Management*.

133. Ludin, S.M. and Arbon, P.A. (2017), "Improving community disaster resilience through scorecard self-testing", Disaster Prevention and Management, Vol. 26 No. 1, pp. 13-27. https://doi.org/10.1108/DPM-08-2016-0177

134. MacTavish, K., and Salamon, S. (2001). "Mobile home park on the prairie: A new rural community form." Rural Sociol., 66(4), 487–506.

135. Malalgoda, C., Amaratunga, D., & Haigh, R. (2014). Challenges in creating a disaster resilient built environment. *Procedia Economics and Finance*, *18*, 736-744. https://doi.org/10.1016/S2212-5671(14)00997-6

136. Margerum, R. D., & Robinson, C. J. (2015). Collaborative partnerships and the challenges for sustainable water management. Current Opinion in Environmental Sustainability, 12, 53-58.10.1016/j.cosust.2014.09.003

137. Martínez-Gomariz, E., Forero-Ortiz, E., Guerrero-Hidalga, M., Castán, S., and Gómez, M. (2020). Flood depth–damage curves for Spanish urban areas. Sustainability, 12(7), 2666.

138. Mathur, V. N. (2007). Defining, identifying and mapping stakeholders in the assessment of urban sustainability. IN: Horner, M.... et al. In *Proceedings: SUE-MoT Conference 2007: International Conference on Whole Life Sustainability and its Assessment, Glasgow, Scotland, 27th-29th June*.

139. McShane M. K. and Wie J.-E., (2019). Toward better management of flood losses: flood insurance in a wetter world. Public Works Management and Policy 24(1), 88–109. https://doi.org/10.1177/1087724X18805500.

140. Miao, Q., & Davlasheridze, M. (2022). Managed retreat in the face of climate change: examining factors influencing buyouts of floodplain properties. *Natural Hazards Review*, *23*(1), 04021063.

141. Michel-Kerjan E. O. & Kunreuther H., (2011). Redesigning flood insurance. Science 333, 408–409. 22 July 2011. Policy Forum. http://science.sciencemag.org/content/333/6041/408

142. 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.

143. Michel-Kerjan, E., Lemoyne de Forges, S., and Kunreuther, H. (2012). Policy tenure under the US national flood insurance program (NFIP). Risk Analysis: An International Journal, 32(4), 644-658.

144. Miguez, M. G., & Veról, A. P. (2017). A catchment scale Integrated Flood Resilience Index to support decision making in urban flood control design. *Environment and Planning B: Urban Analytics and City Science*, *44*(5), 925-946. DOI: https://doi.org/10.1177/0265813516655799

145. Miller, B., Dixon, L., and Clancy, N. (2019). Reasonable and risk-based? Replacing NFIP generally subsidized rates with a means-tested subsidy. Southern Economic Journal, 85(4), 1180-1195.

146. Mobley, W., Sebastian, A., Blessing, R., Highfield, W. E., Stearns, L., and Brody, S. D.: Quantification of continuous flood hazard using random forest classification and flood insurance claims at large spatial scales: a pilot study in southeast Texas, Nat. Hazards Earth Syst. Sci., 21, 807–822, https://doi.org/10.5194/nhess-21-807-2021, 2021.

147. Mochamad, R. F., & Preece, R. (2019). Assessing the Impact of VSC-HVDC on the Interdependence of Power System Dynamic Performance in Uncertain Mixed AC/DC Systems. *IEEE Transactions on Power Systems*, *35*(1), 63-74.

148. Moreira, L. L., de Brito, M. M., & Kobiyama, M. (2021). A systematic review and future prospects of flood vulnerability indices. Natural Hazards and Earth System Sciences, 21(5), 1513-1530.

149. Multi-Hazard Mitigation Council (2019.). Natural Hazard Mitigation Saves: 2019 Report. Principal Investigator Porter, K.; Co-Principal Investigators Dash, N., Huyck, C., Santos, J., Scawthorn, C.; Investigators: Eguchi, M., Eguchi, R., Ghosh., S., Isteita, M., Mickey, K., Rashed, T., Reeder, A.; Schneider, P.; and Yuan, J., Directors, MMC. Investigator Intern: Cohen-Porter, A. National Institute of Building Sciences. Washington, DC. www.nibs.org

150. Munich Re "Natural disaster risks: Losses are trending upwards" (https://www.munichre.com/en/risks/natural-disasters-losses-are-trending-upwards.html#-1624621007) accessed March 14, 2021.

151. Munich Re "The flood insurance gap in the United States" (August 28, 2020) (https://www.munichre.com/topics-online/en/climate-change-and-natural-disasters/natural-disasters/floods/the-flood-insurance-gap-in-the-us.html) accessed March 14, 2021.

152. Murphy. P. (April 1, 2022). Risk Rating 2.0 goes into effect, more than 80% of policy holders could see rate increase. https://www.wwltv.com/article/news/local/risk-rating-20-goes-into-effect-more-than-80-of-policy-holders-could-see-rate-increase/289-6ec30e7c-22ac-4567-95e7-32d42ce695be accessed November 4, 2022

153. National Flood Insurance Program. (2015). Report to Congress on Reinsuring NFIP Insurance Risk and Options for Privatizing the NFIP (Washington, DC: Federal Emergency Management Agency).

154. National Institute of Standards and Technology (NIST). (2020). Community Resilience. Retrieved from: https://www.nist.gov/topics/community-resilience.

155. NOAA National Centers for Environmental Information (NCEI) U.S. Billion-Dollar Weather and Climate Disasters (2021). https://www.ncdc.noaa.gov/billions/, DOI: 10.25921/stkw-7w73

156. North Carolina Department of Public Safety, "83 Lenoir County Homeowners to Receive Buyouts for Properties Flooded by Hurricane Matthew," news release, July 10, 2018, https://www.ncdps.gov/news/press-releases/2018/07/10/83-lenoir-county-homeowners-receive-buyouts-properties-%C2%A0flooded. Accessed July 8, 2023

157. Noy, I., Becerra, O., & Cavallo, E. A. (2012). Foreign Aid in the Aftermath of Large Natural Disasters.Available at SSRN: https://ssrn.com/abstract=2207237 or http://dx.doi.org/10.2139/ssrn.2207237

158. O'Donnell, E., Thorne, C., Ahilan, S., Arthur, S., Birkinshaw, S., Butler, D., ... & Wright, N. (2020). The blue-green path to urban flood resilience. Blue-Green Systems, 2(1), 28-45. doi: 10.2166/bgs.2019.199

159. Oh, E. H., Deshmukh, A., & Hastak, M. (2010). Vulnerability assessment of critical infrastructure, associated industries, and communities during extreme events. In *Construction Research Congress 2010: Innovation for reshaping construction practice* (pp. 449-469).

160. Ouma, Y. O., & Tateishi, R. (2014). Urban flood vulnerability and risk mapping using integrated multi-parametric AHP and GIS: methodological overview and case study assessment. Water, 6(6), 1515-1545.

161. Owusu-Ansah, J. K., Dery, J. M., & Amoako, C. (2019). Flood vulnerability and coping mechanisms around the Weija Dam near Accra, Ghana. GeoJournal, 84(6), 1597-1615.

162. Patankar, A., & Patwardhan, A. (2016). Estimating the uninsured losses due to extreme weather events and implications for informal sector vulnerability: a case study of Mumbai, India. Natural Hazards, 80(1), 285-310.

163. Paul A. Raschky & Hannelore Weck-Hannemann (2007) Charity hazard—A real hazard to natural disaster insurance?, Environmental Hazards, 7:4, 321-329, DOI: 10.1016/j.envhaz.2007.09.002

164. Peng, J., Shan, X. G., Gao, Y., Kesete, Y., Davidson, R. A., Nozick, L. K., & Kruse, J. (2014). Modeling the integrated roles of insurance and retrofit in managing natural disaster risk: A multi-stakeholder perspective. *Natural Hazards*, *74*, 1043-1068

165. Peng, Y., Wang, Z., Cui, Y., Su, Y., Tang, Y., Luo, S., ... & Bai, T. (2022). Prediction of drought/flood intensities based on a 500-year time series in three different climate provinces of China. Regional Environmental Change, 22(2), 80.

166. Penning-Rowsell, E. and Wilson, T.: Gauging the impact of natural hazards: the pattern and cost of emergency response during flood events, T. I. Brit. Geogr., 31(2), 99–115, 2006.

167. Petrolia, D. R., C. E. Landry, and K. H. Coble, 2013, Risk Preferences, Risk Perceptions, and Flood insurance, Land Economics, 89(2): 227-245.

168. Pistrika, A. (2010). Direct flood damage estimation in built environment/ PhD dissertation, School of Rural and Surveying Engineering, National Technical University of Athens, March 2010

169. Pistrika, A. K., and Jonkman, S. N. (2010). Damage to residential buildings due to flooding of New Orleans after hurricane Katrina. Natural Hazards, 54, 413-434.

170. Pistrika, A., Tsakiris, G., and Nalbantis, I. (2014). Flood depth-damage functions for built environment. Environmental Processes, 1, 553-572.

171. Polacek, A. (2018). Catastrophe bonds: A primer and retrospective. *Chicago Fed Letter*, *405*, 1-7.

172. Quinn, N., Bates, P. D., Neal, J., Smith, A., Wing, O., Sampson, C., ... & Heffernan, J. (2019). The spatial dependence of flood hazard and risk in the United States. Water Resources Research, 55(3), 1890-1911.

173. Ramm, T. D., Watson, C. S., & White, C. J. (2018). Strategic adaptation pathway planning to manage sea-level rise and changing coastal flood risk. Environmental Science & Policy, 87, 92-101.

174. Rasch, R. J. (2016). Assessing urban vulnerability to flood hazard in Brazilian municipalities. Environment and urbanization, 28(1), 145-168.

175. Raschky, P. A., Schwarze, R., Schwindt, M., & Zahn, F. (2013). Uncertainty of governmental relief and the crowding out of flood insurance. *Environmental and Resource Economics*, *54*, 179-200.

176. Reguero, B. G., Beck, M. W., Schmid, D., Stadtmüller, D., Raepple, J., Schüssele, S., & Pfliegner, K. (2020). Financing coastal resilience by combining nature-based risk reduction with insurance. *Ecological Economics*, *169*, 106487. https://doi.org/10.1016/j.ecolecon.2019.106487

177. Ren, J., & Wang, H. H. (2016). Rural homeowners' willingness to buy flood insurance. *Emerging Markets Finance and Trade*, *52*(5), 1156-1166.

178. Reuters (2020) Hurricanes Eta, Iota Hit Nicaragua With $743 Million in Economic Losses (https://www.usnews.com/news/world/articles/2020-11-24/hurricanes-eta-iota-hit-nicaragua-with-743-million-in-economic-losses) accessed March 14, 2021

179. Rieger, C. G. (2014, August). Resilient control systems practical metrics basis for defining mission impact. In *2014 7th International Symposium on Resilient Control Systems (ISRCS)* (pp. 1-10). IEEE.

180. Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55.

181. Rozelle, J., Bausch, D., McNabb, S., Robles-Kyle, N., & Region, F. E. M. A. (2010). Analyzing potential flood damage to Fargo, North Dakota using a detailed HAZUS Level 2 analysis.

182. Rui, X., Gongyuan, L., & Dong, H. (2010, June). Simulated annealing algorithm for solving a bi-level optimization model on high-speed railway station location. In *2010 Third International Conference on Information and Computing* (Vol. 2, pp. 159-162). IEEE.

183. Rumbach, A., Sullivan, E., & Makarewicz, C. (2020). Mobile home parks and disasters: understanding risk to the third housing type in the United States. Natural hazards review, 21(2), 05020001.

184. S&P Dow Jones Indices LLC, S&P/Case-Shiller U.S. National Home Price Index [CSUSHPINSA], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CSUSHPINSA, July 27, 2023.

185. Sadler, J. M., Goodall, J. L., Morsy, M. M., and Spencer, K. (2018). Modeling urban coastal flood severity from crowd-sourced flood reports using Poisson regression and Random Forest. Journal of hydrology, 559, 43-55.

186. Sanders, B. F., Schubert, J. E., Goodrich, K. A., Houston, D., Feldman, D. L., Basolo, V., ... and Matthew, R. A. (2020). Collaborative modeling with fine-resolution data enhances flood awareness, minimizes differences in flood perception, and produces actionable flood maps. Earth's Future, 8(1), e2019EF001391.

187. Santos, M., Fragoso, M., & Santos, J. A. (2018). Damaging flood severity assessment in Northern Portugal over more than 150 years (1865–2016). Natural Hazards, 91(3), 983-1002.

188. Scawthorn, C., Blais, N., Seligson, H., Tate, E., Mifflin, E., Thomas, W., ... & Jones, C. (2006a). HAZUS-MH flood loss estimation methodology. I: Overview and flood hazard characterization. *Natural Hazards Review*, *7*(2), 60-71.

189. Scawthorn, C., Flores, P., Blais, N., Seligson, H., Tate, E., Chang, S., ... & Lawrence, M. (2006a). HAZUS-MH flood loss estimation methodology. II. Damage and loss assessment. *Natural Hazards Review*, *7*(2), 72-81. https://doi.org/10.1061/(ASCE)1527-6988(2006)7:2(72)

190. Schauer, B. A., Scawthorn, C., Tertell, P., & Jones, C. (2002). The Hazus Flood Loss Estimation Model for Characterizing Coastal Flood Hazards and Determining Losses. In *Solutions to Coastal Disasters' 02* (pp. 178-185).

191. Scheuer, S., Haase, D., & Meyer, V. (2011). Exploring multicriteria flood vulnerability by integrating economic, social and ecological dimensions of flood risk and coping capacity: from a starting point view towards an end point view of vulnerability. *Natural hazards*, *58*, 731-751.

192. Schilling K., Copeland C., Dixon J., Smythe J., Vincent M. and Peterson J., (1987). The Nation's Public Works: Report on Water Resources. National Council on Public Works Improvement, Washington, DC.

193. Schröter, K., Kreibich, H., Vogel, K., Riggelsen, C., Scherbaum, F., & Merz, B. (2014). How useful are complex flood damage models?. Water Resources Research, 50(4), 3378-3395.

194. Sharma, D. K., Cui, Q., Chen, L., & Lindly, J. K. (2010). Balancing Private and Public Interests in Public–Private Partnership Contracts through Optimization of Equity Capital Structure. Transportation Research Record, 2151(1), 60–66. https://doi.org/10.3141/2151-08

195. Shen, G. (2005). “Location of manufactured housing and its accessibility to community services: A GIS-assisted spatial analysis.” Socioecon. Plann. Sci., 39(1), 25–41.

196. Sheng, J., & Wilson, J. P. (2009). Watershed urbanization and changing flood behavior across the Los Angeles metropolitan region. Natural Hazards, 48, 41-57.

197. Shi, P. (2016). Insurance ratemaking using a copula-based multivariate Tweedie model. Scandinavian Actuarial Journal, 2016(3), 198-215.

198. Shively, D. (2017). Flood risk management in the USA: implications of National Flood Insurance Program changes for social justice. Regional Environmental Change, 17(6), 1663-1672.

199. Shmueli, G. (2010). To explain or to predict? Stat. Sci. 25, 289–310. doi:10.1214/10-sts330

200. Siders, A. R., Hino, M., & Mach, K. J. (2019). The case for strategic and managed climate retreat. Science, 365(6455), 761-763.

201. Spekkers, M. H., Kok, M., Clemens, F. H. L. R., and Ten Veldhuis, J. A. E. (2014). Decision-tree analysis of factors influencing rainfall-related building structure and content damage. Natural hazards and earth system sciences, 14(9), 2531-2547.

202. Stephenson, V., & D'ayala, D. (2014). A new approach to flood vulnerability assessment for historic buildings in England. Natural Hazards and Earth System Sciences, 14(5), 1035-1048.

203. Sterett, S. M. (2015). Disaster, displacement, and casework: Uncertainty and assistance after Hurricane Katrina. Law & Policy, 37(1-2), 61-92.

204. Sterman, J. (2010). *Business dynamics*. Irwin/McGraw-Hill c2000

205. Suriya, S., Mudgal, B. V., & Nelliyat, P. (2012). Flood damage assessment of an urban area in Chennai, India, part I: methodology. Natural Hazards, 62, 149-167.

206. Sylves, R., & Búzás, Z. I. (2007). Presidential disaster declaration decisions, 1953–2003: What influences odds of approval?. State and Local Government Review, 39(1), 3-15.

207. Tarhule, A. (2005). Damaging rainfall and flooding: the other Sahel hazards. Climatic change, 72(3), 355-377.

208. Terti, G., Ruin, I., Anquetin, S., and Gourley, J. J. (2015). Dynamic vulnerability factors for impact-based flash flood prediction. Natural Hazards, 79(3), 1481-1497.

209. Tesselaar, M., Botzen, W. W., Robinson, P. J., Aerts, J. C., & Zhou, F. (2022). Charity hazard and the flood insurance protection gap: An EU scale assessment under climate change. Ecological Economics, 193, 107289.

210. Centre for Research on the Epidemiology of Disasters and United Nations Office for Disaster Risk Reduction. (2015). The Human Cost of Weather Related Disasters, 1995–2015.

211. Thomalla, F., & Larsen, R. K. (2010). Resilience in the context of tsunami early warning systems and community disaster preparedness in the Indian Ocean region. Environmental Hazards, 9(3), 249-265. https://doi.org/10.3763/ehaz.2010.0051

212. Tierney, K. J., M. K. Lindell, and R. W. Perry (2001). Facing the Unexpected: Disaster Preparedness and Response in the United States. Washington, D.C.: Joseph Henry Press.

213. Tsakiris, G. J. N. H. (2014). Flood risk assessment: concepts, modelling, applications. Natural Hazards and Earth System Sciences, 14(5), 1361-1369.

214. U.S. Treasury. (2023). What is the national debt?. https://fiscaldata.treasury.gov/americas-finance-guide/national-debt/ accessed July 2, 2023

215. UNDRR. (2020). The human cost of disasters: an overview of the last 20 years (2000-2019). https://www.undrr.org/publication/human-cost-disasters-overview-last-20-years-2000-2019#:~:text=In%20the%20period%202000%20to,over%20the%20previous%20twenty%20years. Accessed July 5, 2023

216. United Nations (2011). “Fourth United Nations Conference on the Lease developed Countries”. Held in Istanbul, Turkey during May 9-13, 2011.

217. Vaijhala, S., & Rhodes, J. (2015). ReBound - Catastrophe Bonds for Resilience Report.

218. Vaijhala, S., & Rhodes, J. (2018). Resilience Bonds: a business-model for resilient infrastructure. *Field Actions Science Reports*. *The Journal of Field Actions*, (Special Issue 18), 58-63.

219. Wagner, Katherine R H. 2022. "Adaptation and Adverse Selection in Markets for Natural Disaster Insurance." American Economic Journal: Economic Policy, 14 (3): 380-421.

220. Wang, D., Davidson, R. A., Nozick, L. K., Trainor, J. E., & Kruse, J. L. (2020). Computational framework to support government policy-making for hurricane risk management. *Natural hazards review*, *21*(1), 04019012.

221. Wang, Y., & Sebastian, A. (2021). Community flood vulnerability and risk assessment: An empirical predictive modeling approach. Journal of Flood Risk Management, 14(3), e12739.

222. Wang, Y., Li, Z., & Gao, Y. (2015). Minimum revenue guarantee and toll revenue cap optimization for PPP highways: Pareto optimal state approach. *The Baltic Journal of Road and Bridge Engineering*, *10*(4), 365-371.

223. Webber. T. (2017). What Does Being A U.S. Territory Mean For Puerto Rico?. NPR. https://www.npr.org/2017/10/13/557500279/what-does-being-a-u-s-territory-mean-for-puerto-rico accessed October 5, 2023

224. Webster. E. M. (2019) FEMA Individual Assistance Programs: An Overview. Congressional Research Service. https://sgp.fas.org/crs/homesec/R46014.pdf accessed March 22, 2022

225. 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), 1-10.

226. World Bank Group. (2016). *Evaluating Sovereign Disaster Risk Finance Strategies: Case Studies and Guidance*. World Bank.

227. World Bank. (2014). Public-private partnerships : reference guide version 2.0

228. Yabe, T., Rao, P. S. C., & Ukkusuri, S. V. (2021). Resilience of interdependent urban socio-physical systems using large-scale mobility data: Modeling recovery dynamics. *Sustainable Cities and Society*, *75*, 103237.

229. Yabe, T., Rao, P. S. C., Ukkusuri, S. V., & Cutter, S. L. (2022). Toward data-driven, dynamical complex systems approaches to disaster resilience. *Proceedings of the National Academy of Sciences*, *119*(8), e2111997119.

230. Yang, X. L., Ding, J. H., and Hou, H. (2013). Application of a triangular fuzzy AHP approach for flood risk evaluation and response measures analysis. Natural hazards, 68(2), 657-674.

231. Yang, Y., Qian, W., & Zou, H. (2018). Insurance premium prediction via gradient tree-boosted Tweedie compound Poisson models. Journal of Business & Economic Statistics, 36(3), 456-470.

232. Yildirim, E., and Demir, I. (2021). An integrated flood risk assessment and mitigation framework: A case study for middle Cedar River Basin, Iowa, US. International Journal of Disaster Risk Reduction, 56, 102113.

233. Young, A. T. (2008). Replacing incomplete markets with a complete mess: Katrina and the NFIP. International Journal of Social Economics, 35(8), 561-568.

234. Zhang, W., Villarini, G., Vecchi, G. A., & Smith, J. A. (2018). Urbanization exacerbated the rainfall and flooding caused by hurricane Harvey in Houston. Nature, 563(7731), 384-388.

235. Zhang, Y. L., and You, W. J. (2014). Social vulnerability to floods: a case study of Huaihe River Basin. Natural Hazards, 71(3), 2113-2125.

236. Zhang, Z., Zhou, J., Cao, W., & Zhang, J. (2016). Causal inference with a quantitative exposure. Statistical methods in medical research, 25(1), 315-335.

237. Zuzak, C., E. Goodenough, C. Stanton, M. Mowrer, N. Ranalli, D. Kealey, and J. Rozelle. (2021). National Risk Index Technical Documentation. Federal Emergency Management Agency, Washington, DC

VITA

Arkaprabha (Arka) Bhattacharyya
Lyles School of Civil Engineering
Purdue University
550 Stadium Mall Dr.
West Lafayette, IN – 47906

EDUCATION

Doctor of Philosophy, Purdue University, USA (Expected in December 2023)
Civil Engineering/Construction Engineering and Management
Dissertation Title – *"Anatomy of Flood Risk and Flood Insurance in the U.S."*

Graduate Certificate in Systems, Purdue University, USA, 2022
Purdue Systems Collaboratory

Master of Science, Purdue University, USA, 2020
Civil Engineering/Construction Engineering and Management
Thesis Title – *"Framework for Identifying Optimal Risk Reduction Strategies to Minimize the Economic Impacts of Severe Weather Induced Power Outages"*

Bachelor of Engineering, Jadavpur University, India, 2016
Civil Engineering

AWARDS AND RECOGNITIONS

2023 **Trailblazers in Engineering Fellow,** Purdue University

2022 **Crooks Graduate Scholarship**, Lyles School of Civil Engineering, Purdue University

2022 **Professional Development Grant,** Purdue Graduate Student Government

2022 **Conference Travel Grant,** College of Engineering, Purdue University

2022 **Gill Endowment Travel Fund,** Lyles School of Civil Engineering, Purdue University

2021 **Graduate Assistantship,** State Utility Forecasting Group, Purdue University

2021 **Best Presentation** Award, ISEC – 11 Conference, Cairo, 2021

2021 **D.V. Terrell Paper Competition** Winner, American Society of Civil Engineers – Region 4

2021 **Professional Development** Award, Lyles School of Civil Engineering, Purdue University

2021 **Exceptional Teaching and Instructional Support** during the COVID-19 Pandemic, Purdue University Teaching Academy

2021 **Best Research Poster** Award, Midwest Graduate Research Symposium – 2021

2021 **Boiler Changemaker** Award, Purdue University Graduate School

2020 **Crooks Graduate Scholarship**, Lyles School of Civil Engineering, Purdue University

2020 **Best Research Poster** Award, The Purdue Chapter of Sigma Xi

RESEARCH PROJECTS

1. *Anatomy of Flood Risk and Flood Insurance in the U.S.* – (**Dissertation Topic**, Advisor – Dr. Makarand Hastak)
   a. Analysis of the presence of Charity Hazard in the U.S. flood insurance market.
   b. Causal model between flood risk factors and flood insurance payout.
   c. Predictive model for insured flood loss estimation.
2. *Towards a Resilient Electric Power Grid: An Investment Prioritization Decision Framework Integrating Risks of Severe Weather-Induced Outages* (**NSF #1728209**)
   a. Infrastructure interdependence modeling for analyzing cascading impacts of prolonged outages.
   b. Economic impact assessment of severe weather induced prolonged power outages.
   c. Optimal strategy selection framework for grid resilience.
3. *Energy Insecurity and Inequity Analysis* (Research work for **State Utility Forecasting Group at Purdue, supported by Indiana Utility Regulatory Commission**)
   a. Equitable access to energy infrastructure.
   b. Spatial and temporal analysis of energy insecurity.
4. *Health and Vulnerability Assessment of the U.S. Construction Industry*
   a. Forecasting health of the U.S. construction industry using Purdue Index for Construction (Pi-C).
   b. Interindustry vulnerability assessment of the U.S. construction industry.

PUBLICATIONS

- **Peer-Reviewed Journal Papers**

Published

1. **Bhattacharyya, A*.** and Hastak, M. (2023) A Data Driven Approach to Quantify Disparities in Power Outages. *Nature Scientific Reports,* 13(1), 7247. https://doi.org/10.1038/s41598-023-34186-9

2. **Bhattacharyya, A*.**, Morshedi, M., and Hastak, M. (2023). A Clustering-Classification Approach in Categorizing Vulnerability of Roads and Bridges using Public Assistance Big Data. *International Journal of Disaster Risk Reduction*,103448. https://doi.org/10.1016/j.ijdrr.2022.103448
3. Morshedi, M., Yoon, S., **Bhattacharyya, A.**, Jung, J., and Hastak, M. (2023). Engaging Engineering Students with the Stakeholders for Infrastructure Planning. *Buildings*, *13*(1), 39. https://doi.org/10.3390/buildings13010039
4. **Bhattacharyya, A.*** and Hastak, M. (2022). Indirect Cost Estimation of Winter Storm Induced Power Outage in Texas. *Journal of Management in Engineering*, https://doi.org/10.1061/(ASCE)ME.1943-5479.0001084 (**Featured in ASCE's Special Collection on Energy Grid and Extreme Weather**)
5. Jeon, J., Padhye, S., **Bhattacharyya, A.**, Cai, H., and Hastak, M. (2022). Impact of COVID-19 on the U.S. Construction Industry as Revealed in the Purdue Index for Construction (Pi-C). *Journal of Management in Engineering*, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000995
6. **Bhattacharyya, A.**, Yoon, S., and Hastak, M. (2021). Economic Impact Assessment of Severe Weather–Induced Power Outages in the US. *Journal of Infrastructure Systems*, 27(4), 04021038. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000648
7. **Bhattacharyya, A.**, Yoon, S., and Hastak, M. (2021). Optimal Strategy Selection Framework for Minimizing the Economic Impacts of Severe Weather Induced Power Outages. *International Journal of Disaster Risk Reduction*, 102265. https://doi.org/10.1016/j.ijdrr.2021.102265
8. **Bhattacharyya, A.**, Yoon, S., Weidner, T. J., and Hastak, M. (2021). Purdue Index for Construction Analytics: Prediction and Forecasting Model Development. *Journal of Management in Engineering*, 37(5), 04021052. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000944

Under Review

1. **Bhattacharyya, A.*,** Wang, H.H., and Hastak, M. Impact of Post-Disaster Federal Support on the National Flood Insurance Program. *International Journal of Disaster Risk Reduction*
2. **Bhattacharyya, A**.* and Hastak, M. Empirical Model for Quantifying the Causal Relations between Flood Risk Factors and Flood Insurance Payout in the U.S. *Journal of Environmental Management*
3. **Bhattacharyya, A**.*, Wang, H. H., and Hastak, M. Regression Model for Predicting Insured Flood Loss to Households in the U.S. *Natural Hazards Review*
4. **Bhattacharyya, A**.* and Hastak, M. Intersectoral Vulnerabilities: Assessing the Impact of External Shocks on the U.S. Construction Sector. *Journal of Management in Engineering*

* Corresponding Author

- **Conference Papers**

Published

1. **Bhattacharyya, A.**, Morshedi, M., and Hastak, M. (2024). Prioritizing Capacity Building Strategies to Ensure Robustness and Faster Recovery of Cellular Networks from Hurricanes. ASCE Construction Research Congress 2024 (Accepted for Publication)
2. Morshedi, M., **Bhattacharyya, A.**, and Hastak, M. (2024). Exploring Socio-Demographic Inequalities in Post-Disaster Community Well-Being: Case Study of Hurricane Harvey. ASCE Construction Research Congress 2024 (Accepted for Publication)
3. **Bhattacharyya, A.**, Wang, H.H., and Hastak, M. (2023). Role of Post-Disaster Federal Payouts on Flood Insurance in the U.S. ASCE INSPIRE 2023. Washington D.C., USA (Accepted for Publication)
4. **Bhattacharyya, A.** and Hastak, M. (2022) Analysis of Federal Expenses to Restore, Repair, Reconstruct, or Replace Disaster Damaged Roads and Bridges in the U.S. 9th International Conference on Construction Engineering and Project Management 2022, Las Vegas, Nevada, USA
5. Morshedi, M., **Bhattacharyya, A.** and Hastak, M. (2022) Explaining the Changes in the Greenhouse Gas Emissions of New York City Buildings. World Building Congress 2022, Melbourne, Australia
6. **Bhattacharyya, A.** and Hastak, M. (2022) Economic Vulnerability Assessment of the Construction Industry in the United States. In Proceedings, Construction Research Congress 2022 https://doi.org/10.1061/9780784483978.027 , Washington D.C., USA
7. **Bhattacharyya, A.**, and Hastak, M. (2022). Impact Analysis of Covid-19 Pandemic on Construction Employment in the United States. In Canadian Society of Civil Engineering Annual Conference (pp. 411-420). Springer, Singapore.
8. **Bhattacharyya, A.** and Hastak, M. (2021) ‘Framework for Planning Capacities to Maximize the Feasible Resilience of Network Infrastructure Systems’ *Proceedings of International Structural Engineering and Construction*, El-Baradei, S., Abodonya, A., Singh, A. and Yazdani, S. (eds.), Vol. 8, Issue 1, ISEC Press, 2021. www.doi.org/10.14455/ISEC.2021.8(1).INF-05, Cairo, Egypt **(Best Presentation Award)**
9. **Bhattacharyya, A.** and Hastak, M. (2021). 'Strategic Investment in Infrastructure Development to Facilitate Economic Growth in the United States'. World Academy of Science, Engineering and Technology, Open Science Index 174, International Journal of Economics and Management Engineering, 15(6), 600 - 605.

- **Thesis and Dissertation**

1. Bhattacharyya, A. Anatomy of Flood Risk and Flood Insurance in the U.S. (Expected in December 2023)
2. Bhattacharyya, A. (2020). Framework for Identifying Optimal Risk Reduction Strategies to Minimize the Economic Impacts of Severe Weather Induced Power Outages. Purdue University Graduate School. Thesis. https://doi.org/10.25394/PGS.12733412.v1

- **Poster Presentation**

1. **Bhattacharyya, A.** and Hastak, M. (2022) A Clustering-Regression Approach in Analyzing Infrastructure Vulnerability Using Public Assistance Big Data. Purdue University Civil Engineering Graduate Student Advisory Council Annual Research Symposium 2022.
2. **Bhattacharyya, A.** and Hastak, M. (2021) Natural Disaster Insurance for U.S. Infrastructures: Good Idea or Requirement. Society of Risk Analysis Annual Meeting 2021
3. **Bhattacharyya, A.**, Morshedi, M., Hastak, M. (2020) Multivariate Prediction Model for Greenhouse Gas Emission of Buildings in New York City. Society of Risk Analysis Annual Meeting 2020
4. **Bhattacharyya, A.**, Yoon, S., Hastak, M. (2020) Economic Impact Assessment due to Severe Weather Induced Power Outages in the U.S. Graduate Student and Post-Doctoral Fellows Research Poster Competition 2020, The Purdue Chapter of Sigma Xi, the Scientific Research Honor Society (**Best Poster Awards**)

PROPOSAL WRITING

1. "Towards Flood Resilient Cities: An Integrated Multistakeholder Collaboration Framework for Planning and Financing Urban Flood Resilience" – (PI – Dr. Makarand Hastak, Co-PI – Dr. Venkatesh Merwade, and Dr. Holly H. Wang) – **Under Preparation**
2. "Analyzing the Nexus of Flood Insurance, Collective Disaster Memory, and Evacuation Behavior based on Evidence from Hurricane Ida", **NSF – RAPID** (2021) - **Declined**

INVITED TALKS

1. "Towards Weather Resilient Electricity Grid - A Decision Making Framework for Utilities", Purdue University, Civil Engineering Graduate Student Advisory Council, Research Bytes Fall 2021
2. "Reducing Economic Impacts of Severe Weather Induced Power Outages", One Concern, Inc, February 2022
3. "Towards Disaster Resilient Electricity Grid: Risk Quantification and Optimal Strategy Planning", Purdue University, CEM-CIB Seminar Series, Fall 2023

TEACHING EXPERIENCE

Graduate Teaching Assistant, Civil Engineering and Construction Engineering and Management, Purdue University

1. Infrastructure Planning – CEM 530 (Graduate Course)
2. Construction Business Management – CE 521 (Graduate Course)
3. Leadership and Advanced Project Management (CEM 497/ CE 497/ CE 597)
4. Legal Aspects of Engineering – CEM 485

Invited Lectures, Purdue University

1. *Course:* ***CEM 530 – Infrastructure Planning****, Lecture on Financial Analysis, Spring 2020*
   - The lecture included present value analysis, future value analysis, equivalent annual worth analysis, benefit-cost analysis, internal rate of return calculation, etc.

2. *Course:* ***CEM 530 – Infrastructure Planning****, Lecture on Time Series Forecasting, Spring 2020*

   - The lecture included statistical techniques such as Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) on R for analyzing popular construction relevant indexes such as Construction Backlog Indicator, Construction Employment, etc.

3. *Course:* ***CEM 530 – Infrastructure Planning****, Lecture on Input-output Model, Spring 2020*

   - The lecture introduced a popular econometric analysis tool named "Input-output Model" for understanding the relationships between different producing and consuming sectors in the U.S. economy. Importance was given explaining how the U.S. construction industry is dependent on other industries and how a shock in other industries can cascade through the economy to influence the U.S. construction industry.

4. *Course:* ***CEM 530 – Infrastructure Planning****, Lecture on Infrastructure Interdependence Modeling, Spring 2023*

   - The lecture explained how infrastructure works a system of systems and how we can model their interdependence following different techniques such as system dynamics simulation, graph theory, input-output model, etc.

WORK EXPERIENCE

Senior Civil Engineer, Larsen and Toubro Limited, India, 2016 – 2018

- Worked in a construction project of 4.5 miles long bridge in Kolkata, India.
- Supervised construction of over 500 bored and cast-in-place pile foundation.
- Optimized project schedules to ensure timely completion and planned profitability.

DIVERSITY EXPERIENCE

Attended the ADVANCE/OVPEC Faculty Search Committee workshop at Purdue in 2019. The goal of this workshop is to increase search committee members' knowledge about current search and hire best practices and procedures leading to the employ of an excellent and diverse faculty.

PROFESSIONAL DEVELOPMENT

- **Certifications**
  1. Graduate Certificate in Systems, Purdue Systems Collaboratory, 2022
  2. Applied Management Principles, Krannert Executive Education Program, Purdue University, 2021
  3. Nature-based Solutions for Disaster and Climate Resilience, UN Environment Program, 2021
- **Faculty Training Workshop/Conference**
  1. 13th Annual Conference for Assistant Professors "Persistence and Resilience: Envisioning what Institutions can do for Faculty", Purdue University, 2022
  2. Trailblazers in Engineering Workshop, Purdue University, 2023
- **Affiliations**
  1. American Society of Civil Engineers, Student Member, 2020 – Present
  2. Society of Risk Analysis, Student Member, 2020 – Present
  3. Buried Asset Management Institute – International (BAMI-I), Member, 2022 – Present
- **Reviewer for Conferences/Journals**
  1. The 11th International Structural Engineering and Construction Conference, 2021
  2. Construction Research Congress, 2022, Washington D.C., USA
  3. Construction Research Congress, 2024, Des Moines, Iowa, USA
  4. The 9th International Conference on Construction Engineering and Project Management, 2022
  5. ASCE Journal of Infrastructure Systems
  6. ASCE Journal of Management in Engineering
  7. International Journal of Mass Emergencies and Disasters
  8. International Journal of Disaster Risk Reduction
  9. Discover Sustainability by Nature
  10. Journal of Applied Economics

- **Other Services**

1. **Graduate Student Mentoring**: Participating in Civil Engineering Graduate Student Advisory Council's (CEGSAC) graduate student mentoring program to help the new coming graduate students to navigate graduate life. Students mentored:
    - Prakriti Suri (Master of Science in Civil Engineering)
    - Priyanka Venkatesa Palanichamy (Master of Science in Civil Engineering)
    - Hyewon Seo (Ph.D. in Civil Engineering)
    - Ikechukwu Onuchukwu (Ph.D. in Civil Engineering)
2. **Graduate Student Coordinator**: Assisted Purdue team at the Global Leadership Forum for Construction Engineering and Management (GLF – CEM) between July 2020 and September 2021. Helped in organizing the group meetings, preparing meeting minutes, updating, and maintaining the GLF – CEM website.
3. **Representative Graduate Student**: Served as the Representative Graduate Student in the CEM/CE Faculty Search Committee (2019) to help to identify an excellent talent pool and recruit the best candidate for the position to Purdue Engineering.
4. **Placement Coordinator**: Represented the School of Civil Engineering at Jadavpur University, India as the department coordinator at the university placement cell to facilitate the placement process for the graduating students.
    - maintaining student and alumni database,
    - communicating with companies that were interested in recruiting students, inviting them to the university campus,
    - organizing logistic support and facilitating the process.