A Data-intensive Approach to Allocating Owner vs. NFIP portion of Average Annual Flood Losses

Md Adilur Rahim[1], Carol Freidland[1], Robert Rohli[1], Nazla Bushra[1], and Rubayet Bin Mostafiz[1]

[1]Louisiana State University

November 25, 2022

Abstract

Accurate loss assessment plays a vital role in understanding the economic risk of natural hazards, for planning, mitigation, and actuarial purposes. Because of its juggernaut status as the most widespread and costly hazard, both nationally and around the world, loss assessment due to flood is particularly important. One of the shortcomings in existing flood loss models is to partition the structure (or building) economic value of loss into that borne by the homeowner and that covered by flood insurance. The goal of this research is to model the loss incurred by the homeowner and that incurred by the National Flood Insurance Program, considering flood damage, building replacement value, flood insurance coverage amount, deductible, and flood characteristics (slope and y-intercept of the loss vs. return period curve). A Monte Carlo approach is used to calculate the annual average loss due to flood at the individual homeowner scale. Multiple linear regression (MLR) and Classification and Regression Tree (CART) models are trained to provide the output of the owner's share of the loss. The CART model outperformed the MLR model with lower RMSE and MSE values and a higher R^2 value (0.95) on the test data set. Because out-of-pocket expenses due to flood can be devastating to financial security, the results of this study support and inform the proactive decision-making process that homeowners can use to self-assess their degree of preparation and vulnerability to the flood hazard.

A Data-intensive Approach to Allocating Owner vs. NFIP portion of Average Annual Flood Losses

Md Adilur Rahim*, Carol J. Friedland, Robert V. Rohli, Nazla Bushra, Rubayet Bin Mostafiz
Louisiana State University

Introduction

Flooding is the costliest natural hazard globally and nationally, in terms of loss of life and property, with impacts felt disproportionately by the economically disadvantaged. Quantification of total flood losses is important for monitoring and mitigating the flood hazard across space and time. In light of the ever-increasing value of property at risk, policymakers are increasingly adopting the approach of integrated flood risk management, which includes the engagement of households in flood insurance and structural flood protection measures at the micro-level. Existing research tends to emphasize quantification of total loss rather than the direct economic impact on owners; it does not explicitly identify the financial contribution of the National Flood Insurance Program (NFIP) relative to that covered by the homeowner, at the individual flood-insured residence scale.

This study partitions the total modeled residential building flood damage between out-of-pocket expense to the owner and that covered by the National Flood Insurance Program (NFIP). Input variables include the Gumbel extreme-value distribution parameters (i.e., the slope (α) and y-intercept (u) of the flood depth as a function of return period), first floor elevation above the adjacent ground surface, building value (i.e., the product of construction cost and livable space), and flood insurance coverage and deductible. Then, Monte Carlo simulation is performed to partition the AAL between the owner and that reimbursed by the NFIP. The contribution of this paper is to help the individual homeowner to make a more informed decision regarding the purchase of flood insurance coverage and the associated deductible. Results from this work can be incorporated into webtools or other education/outreach material made available to the general public and to realtors, homebuilders, and community leaders.

This research takes the next step forward by filling the research gap regarding the allocation of owner and NFIP shares of residential flood loss. Thus, these results will be of interest to homeowners, insurance companies, and lending institutions.

Methods

Input Parameters

The parameters used to run Monte Carlo simulation to model the average annual loss (AAL) are described below:

Flood Hazard Parameters

Gumbel parameters (α, u) that vary such that $0<\alpha<10$ and $-20< u <20$, with α increasing in increments of 2 feet and u increasing in increments of 4 feet. These ranges of α and u are selected because they allow for the 100-year flood depth to range from -20 to 66 feet; values that would encompass the entire range that could be experienced in the U.S.A.

First Flood Elevation

The first-floor elevation (FFE; i.e., above the adjacent surface) is calculated based on the assumption that it could range from 0 feet (i.e., the base flood elevation, BFE) to a height of 3.5 feet above the BFE. This additional elevation, known as freeboard, is computed at increments of 0.5 feet.

For each of the α and u combinations, the BFE is computed based on the 100-year flood depth, which is typically used as the BFE across the U.S.A. ($BFE = u - \alpha * \ln[-\ln[1-(1/100)]]$).

Building Value

Building value (i.e., replacement value) is defined as the product of the livable area and the per unit construction/repair cost of the structure. The building values are selected such that livable areas range from 500 sqft to 5000 sqft, and repair costs range from 50 to 300 USD per sqft.

Insurance Coverage and Deductible

According to Flood Insurance Manual Rates (2021), the minimum deductible is $1,000 if the building coverage is less than or equal to $100,000 and $1,250 if building coverage is over $100,000, with the maximum deductible as $10,000 regardless of building coverage. The deductibles are selected based on this minimum and maximum values. The selected flood insurance (NFIP) coverage ranges from $25,000 to $250,000.

Average Annual Loss Allocation

Decision Rules

Four decision rules are set up to allocate the AAL between the owner and the NFIP. Specifically, 1) if the AAL is less than the insurance coverage, then the owner portion of the loss is considered to be the deductible, for all cases in which the AAL exceeds the deductible; 2) If the AAL does not exceed the deductible, then the owner suffers the entire loss and NFIP's share is zero; 3) As long as the AAL is less than the coverage, the NFIP portion is the difference between the AAL and the owner's portion; 4) If the AAL exceeds the insurance coverage, then the owner's portion of the loss is equal to the difference between the AAL and the coverage, plus the deductible, and NFIP's share of the loss is the coverage minus the deductible (or AAL minus owner's share).

Monte Carlo Simulation

Monte Carlo simulation of 50,000 flood events (i, where i increases iteratively from 1 to 50,000) is then run for each combination of α, u, FFE, building value, coverage, and deductible. The Monte Carlo simulation provides a random value between 0 and 1 for each of the i runs, in order to generate an exceedance probability (P_i) such that:

P_i = random(0,1)

Then, the flood elevation for each simulated event (FE_i) is calculated using α, u, and P(i);

$FE_i = u - \alpha * \ln P_i$

The flood elevation above the FFE is then calculated as $FE_i - FFE$. The function expressing the annual average loss to the structure is determined by the following equations, implemented by Gnan (2021):

$$AAL_i (\%) = \frac{0.0092\,(FE_i-FFE)^3 - 0.5342\,(FE_i-FFE)^2 + 1\,.404\,(FE_i-FFE) + 13.418}{100}\quad ; \text{where } FE_i - FFE > -1$$

$AAL_i (\%) = 0$; where $FE_i - FFE \leq -1$

The dollar value of the average annual loss is calculated using the following equation:

$AAL_i = AAL_i (\%) *$ Building Value ($)

At this point, the values for coverage and deductible in the scenario under consideration are input so that AAL_i can be partitioned into that cost borne by the owner vs. that assigned to the NFIP, as described by the four decision rules above. The values from these 50,000 runs are then averaged to calculate the AAL, and the portion of the loss assigned to the owner and to the NFIP. The owner ratio is then calculated as the owner share divided by the total AAL.

Result

Data Generation:

A data set is generated using the Monte Carlo method for all the combination of input parameters. A small sample for the owner share, NFIP share, and ratio of owner share to the total, for a few combinations of input parameters, is shown in Table 1.

A test data set is also generated using the parameter values that are not present in the training/validation data set. The data set is then tested to determine the extent to which "intermediate" values for α(1.5, 3.3, 4.7, 8.2), u (-4.86, -1.39, 2.77, 5.89), FFE (BFE + 0.7, BFE + 1.33, and BFE + 2.88), building value (137,500 and 215,000), insurance coverage (45,500, 157,500, and 212,950), and deductible (1,250 and 3,300) fit well with the trained model.

Regression Analysis

The generated data set is used to train regression models that can predict the ratio of owner loss to AAL with reasonable accuracy. The scikit-learn Python library is used to preprocess and train the models. The data set is divided between training (80 percent of the observations) and validation (20 percent) components. The "train_test_split" method in the "sklearn.model_selection" package is used for splitting the data set. The MinMax scaler is used to scale the variables. The "GridSearchCV" algorithm in the scikitlearn.model_selection package is utilized to tune the models hyperparameters. The test data set is used to check the performance of the regression models.

Table 1. Sample scenario values.

α	u	FFE	Building value	Coverage ($)	Deductible ($)	Owner Share ($)	NFIP Share ($)	Ratio (Owner/Total)
6	-8	20.1	200,000	125,000	1,000	97.5	849.3	0.103
6	-8	20.1	200,000	125,000	1,500	111.3	880.0	0.112
6	-8	20.1	200,000	125,000	2,000	96.7	852.0	0.102
6	-8	20.1	200,000	125,000	2,500	106.6	842.1	0.112
6	-8	20.1	200,000	125,000	3,000	102.7	849.7	0.108
6	-8	20.1	200,000	125,000	5,000	132.9	808.2	0.141
6	-8	20.1	200,000	125,000	10,000	186.7	759.2	0.197
2	4	14.7	250,000	150,000	1,000	11.1	405.1	0.027
2	4	14.7	250,000	150,000	1,500	15.6	408.2	0.037
2	4	14.7	250,000	150,000	2,000	18.5	427.9	0.042
2	4	14.7	250,000	150,000	2,500	21.4	374.2	0.054
2	4	14.7	250,000	150,000	3,000	26.5	409.5	0.061
2	4	14.7	250,000	150,000	5,000	42.3	377.6	0.101
2	4	14.7	250,000	150,000	10,000	84.8	352.2	0.194

A comparison of the model fit diagnostics for the validation and test data sets reveals that CART performs more strongly than multiple linear regression (Table 2). Thus, no further complex models are tested, and CART is selected as the final model with 0.95 RMSE value at the test dataset.

Table 2. Model fit diagnostics for predicting the owner share as a proportion of the annual average loss, using the minmax scaling approach.

	Validation Data Set		Test Data Set	
	Multiple Linear Regression	CART	Multiple Linear Regression	CART
Mean squared error	0.0269	0.0007	0.0089	0.0017
R-square	0.4896	0.9868	0.7463	0.9504
Root-mean-square error	0.1639	0.0264	0.0944	0.0417

Summary & Conclusion

This study allows for the development of an interactive decision-making tool in which the user could input the building value , FFE, amount of flood insurance coverage, and flood insurance deductible for a new construction. The output would then provide an estimate of the amount that the user should be prepared to incur in out-of-pocket expenses based on the probabilities of flood likelihood for the home in question and the insurance parameters provided.

The tool would support a more thoughtful selection of flood insurance coverage for the homeowner, to optimize expenses (either through elevating the home, increasing insurance coverage, or both) in consideration of the local hazard. Similar benefits would be available to community planners and leaders as the results are upscaled from the individual to the community level, as they consider options to enact policy and/or infrastructure to mitigate the hazard.

It should be noted that in this analysis, only the direct losses to the structure (i.e., home) are considered, such as the removal and replacement of flooring and drywall; losses to building contents, such as furniture, vehicles, clothing, and items of sentimental value, and indirect losses, such as time unemployed and hotel expenses incurred during renovation are not considered in this analysis. Furthermore, for convenience and ease of comparison across parameters and structures, all costs and payouts are expressed in terms of average annual values, even though it is recognized that the costs will actually be zero in most years and much greater than the mean value following a flood. In future work, the methodology can be expanded to include some or all of these indirect losses, if the relevant loss functions are known. Likewise, content loss (i.e., monetary, but not sentimental, losses to the goods within the structure) can be included in future research if loss functions are known.

LSU