MODELING METHODOLOGY

FEB 16, 2021

Authors

Shirish Chinchalkar – Mng. Director, Predictive Analytics

Pouyan Mashayekh – Sr. Director, Predictive Analytics

Contact Us

Americas
+1.212.553.1658
clientservices@moodys.com

Europe
+44.20.7772.5454
clientservices.emea@moodys.com

Asia (Excluding Japan)
+85 2 2916 1121
clientservices.asia@moodys.com

Japan
+81 3 5408 4100
clientservices.japan@moodys.com

Mortgage Portfolio Analyzer:

Capturing The Impact of Hurricanes and Floods on US Mortgage Defaults Using 427 Scores

Abstract

Hurricanes and floods cause mortgage borrowers to default. Mortgage lenders, mortgage insurers, and financial regulators need to know the impact of these events on mortgage portfolio losses. In this paper, we measure the impact of hurricanes and floods on mortgage

losses using loan-level mortgage data, 427 scores[1], and a Federal Emergency Management Agency (FEMA) data set that contains more than 13,000 natural disasters in the US. We successfully quantify the impact of these events on mortgage defaults and 427 scores are used to capture intra state variability. Moreover, using the Maximum Distribution Theorem, we estimate the probability of severe events at the state level. The models were used to show the impact of 427 scores on mortgage PDs.

[1] Four Twenty Seven is a California-based company that calculates scores for different climate related hazards at the address level.

Table of Contents

1. Introduction

Natural disasters cause borrowers to cease loan payments due to financial difficulties like the loss of jobs or by the destruction of their assets such as houses, cars, and small businesses. Lenders usually implement lenient policies to accommodate borrowers when natural disasters happen. However, many borrowers may still default because of financial shocks emanating from these natural disasters. In this paper, the study is limited to the impact of hurricanes and floods on mortgages.

Measuring the impact of natural disasters on consumer credits, especially mortgages, has been a challenging task for modelers. The events are few and their impacts are local. In this research paper we observe and estimate the impact of hurricanes and floods on mortgages. In Section 2, we explain how those natural disasters have impacted the probability of default (PD) of a mortgage. In Section 3, we show how models are derived from FEMA data to simulate natural disasters at the state level. In Section 4, 427 scores are explained, and we show how they can be used to capture the intra state variability of the impacts. Section 5 shows the empirical results.

2. The impact of natural disasters on the PD in the historical data set

For this study, we used a large loan level data set of mortgages from Freddie Mac and Fannie Mae. Monthly performance of these mortgages was monitored from 2004 to 2020 and a model for the default probability was built using a logistic regression. Details of this model can be found elsewhere.[2]

The US experiences hurricanes and floods frequently. Hurricanes like Sandy, Katrina, and Irma had adverse impacts on southern states. Borrower behaviors were affected by those events. As an example, Figure 1 shows the calendar time series of actual and predicted mortgage defaults in Texas. Mortgage default is defined as the first time a loan is 90 days past due. There is a sharp increase in defaults at the end of 2017 which is caused by hurricane Harvey.

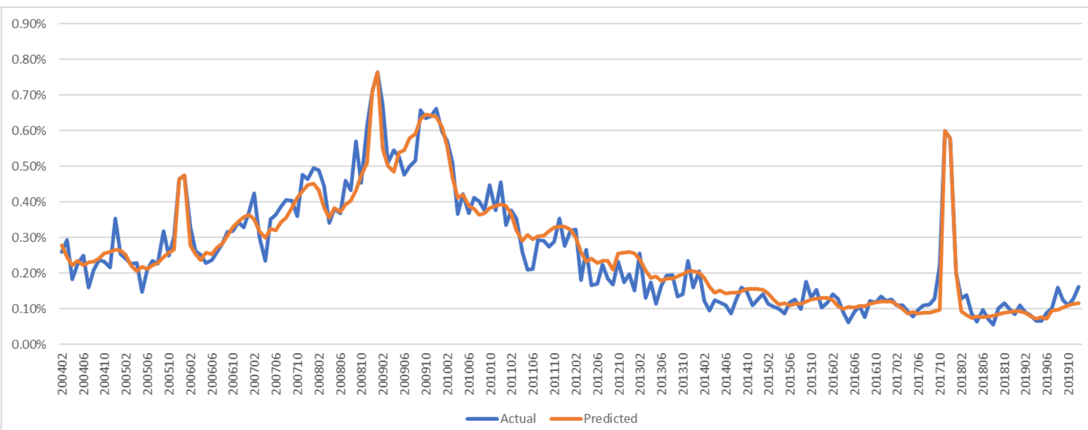

Figure 1. Monthly probability of default for mortgages in Texas

A similar pattern has been observed for hurricane Katrina in 2005 in Alabama and Mississippi. Figures 2 and 3 show the monthly default rate for mortgages in Alabama and Mississippi respectively.

[2] Mortgage Portfolio Analyzer – A model for mortgage portfolio losses, Shirish Chinchalkar, Pouyan Mashayekh, Jui-Chuan Wu, Bahar Kartalciklar, Moody's Analytics, February 2021.

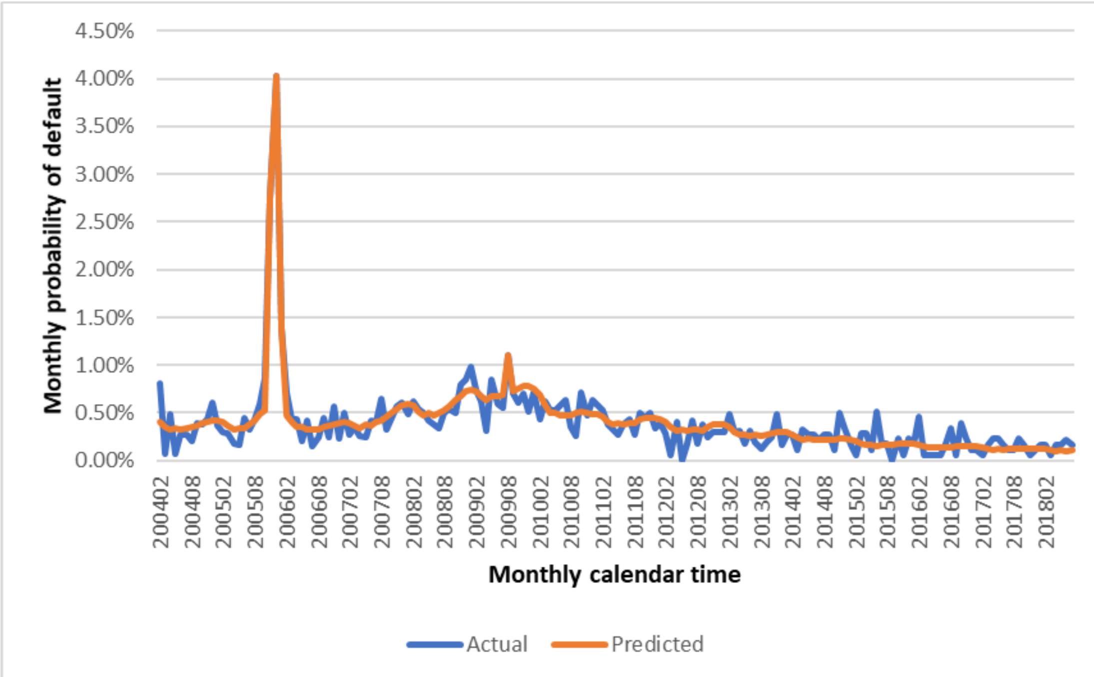

Figure 2. Monthly probability of default for mortgages in Mississippi

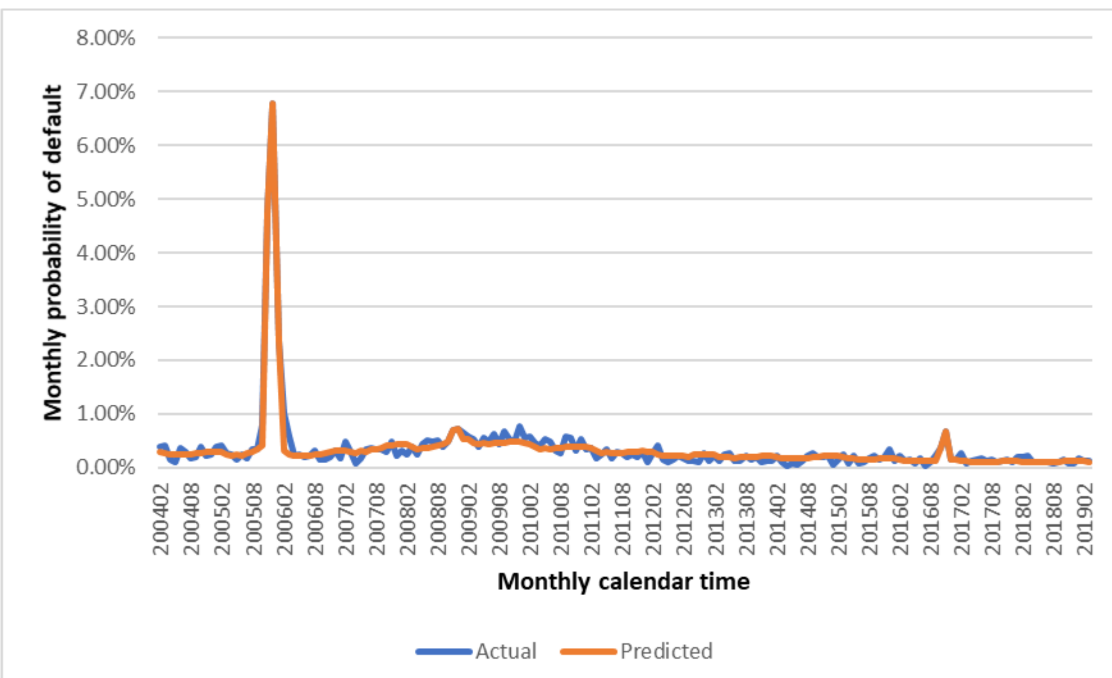

Figure 3. Monthly probability of default for mortgages in Alabama

The spikes in mortgage defaults during Katrina are obvious in both States.

To capture the impact of natural disasters, state-specific calendar time dummies were used in the default model. Since the dataset is very large, a 10 percent random sample was used for estimation. To capture the impact of hurricanes and floods, another dataset was created by limiting the datapoints to those states and months in which mortgages were affected by hurricanes and floods. Since the data was limited to only those months and states, the entire sample was used for the second regression (more than 2.2 million observations). After removing the impact of the climate related dummy variables in the first regression, 427 scores were used to explain the gap between actual default probabilities and the predicted probabilities from the first regression. the average probabilities of actual default and predicted ones in the second dataset are 36 and 7 basis points respectively. Table 1 shows the list of states and months included in the second dataset.

Table 1. List of different State and months selected for

State	Year and Month
AL	2005M12,2006M01
FL	2004M12,2017M01,2017M12,2018M01
LA	2005M11,2005M12,2006M01,2008M11,2008M12,2016M10,2016M11
MS	2005M11,2005M12,2006M01,2009M08
NC	2018M11,2018M12
NJ	2011M09,2012M12,2013M01,2013M02,2013M03,2013M04
OK	2008M08
PR	2017M11,2017M12,2018M01,2018M02,2018M03,2018M04
TN	2012M05
TX	2005M12,2006M01,2008M12,2009M01,2017M11,2017M12,2018M01

In this research, the analysis is limited to flood, hurricane, tornado, thunderstorm, and heavy rain. Table 2 shows the top eight costliest events.

Table 2. List of the top eight costliest hurricanes and floods

state	date	damage property in $	deaths	disaster
TX	201708	43,704,268,000	67	Harvey
NJ	201210	20,950,000,000	2	Sandy
PR	201709	19,018,177,000	20	Irma
LA	200508	16,933,030,000	816	Katrina
MS	200508	13,482,120,000	181	Katrina
FL	200409	10,562,815,000	13	Stewart
FL	200510	10,215,603,000	1	Wilma
LA	201608	8,992,219,000	12	32 inches of rainfall

3. Modeling natural disasters

To simulate extreme events, the probability distribution of the events is needed. A disadvantage of non-parametric estimation is the low frequency of observations in the tail of the distribution. This leads to estimates that exhibit high volatility, especially in the tail of the distribution, which is important for simulating extreme events. The theory of statistical extreme value mitigates the problem by introducing a parametric distribution function for the tail. Extreme Value theory has emerged as one of the most important statistical disciplines for the applied sciences over the last few decades. Extreme value techniques are also becoming widely used in many other disciplines. The distinguishing feature of extreme value analysis is the objective to quantify the stochastic behavior of a process at unusually large – or small – levels. In particular, extreme value analyses usually require estimation of the probability of events that are more extreme than any that have already been observed.

It is natural to regard extreme events as those observations of a variable, X_i, that exceed some high threshold, u. Denoting an arbitrary term in the X_i sequence by X, it follows that a description of the stochastic behavior of extreme events is given by the conditional probability:

$$Pr\{X > u + y | X > u\} = \frac{1-F(u+y)}{1-F(u)}, y > 0 \quad (1.1)$$

If the parent distribution, F, were known, the distribution of threshold exceedances would be known. Since, in practical applications, this is not the case, approximations that are broadly applicable for high values of the threshold are sought. This parallels the use of the Generalized Extreme Value (GEV) as an approximation to the distribution of maxima long sequences when the parent population is unknown. Let $X_1,\ldots,X_n$ be a sequence of independent random variables with common distribution function F. Then for large enough u, the distribution function of $(X - u)$, conditional on $X > u$ is approximately:

$$H(y) = 1 - (1 + \frac{\theta . y}{\bar{\sigma}})^{-\frac{1}{\theta}}$$

Defined on $\{ y : y > 0$ and $(1 + \frac{\theta y}{\bar{\sigma}}) > 0\}$ where $\bar{\sigma} = \sigma + \theta(u - \mu)$

The family of distributions defined by (1.1) is called the generalized Pareto family. So threshold excesses have a corresponding approximate distribution within the generalized Pareto family. Having determined a threshold, the parameters of the generalized Pareto distribution can be estimated by maximum likelihood. Suppose that the values $Y_1,\ldots,Y_n$ are the K excesses of a threshold u. For $\theta \neq 0$ the log-likelihood is derived from (1.1) as:

$$l(\sigma,\theta) = -KLog(\sigma) - (1 + 1/\theta) \textstyle\sum_{i=1}^{K} (1 + \frac{\theta y_i}{\sigma})^{\frac{-1}{\theta}} \quad (1.2)$$

for $i = 1, \ldots, K$ otherwise $l(\sigma,\theta) = -\infty$.

In the case $\theta = 0$ the log-likelihood is:

$$l(\sigma) = -KLog(\sigma) - \sigma^{-1} \sum_{i=1}^{K} y_i$$

The distribution will be simplified to an exponential distribution.
The probability distribution of each event is estimated at the state level. There are two parameters estimated for every state for each of two events, the probability of occurrence and the severity. As an example, in Florida, the monthly probabilities of hurricanes and wildfires are 2.9% and 0.09% respectively. The severities for the two events for Florida follow exponential distributions with parameters estimated independently. These events are simulated and added to the PD model.

4. 427 scores and model estimation

Four Twenty Seven's Sub-Sovereign Physical Climate Risk Scores bring together physical hazard and population data to assess exposure to climate change at various sub-sovereign administrative boundaries. The sub-sovereign scores are based on an index scoring method that draws on a spatially-explicit representation of population exposure to climate hazards. Across the world's land area, 19 indicators for six hazards—floods, heat stress, hurricanes and typhoons, sea level rise, water stress, and wildfires— were evaluated and intersected with projected population in 2040. Based on this analysis, sub-sovereign entities a score reflecting levels of climate risk exposure relative to the other entities of the same type in the country or region of interest (e.g., counties in the United States, NUTS1-3 in the European Union, etc.).

All hazard scores are calculated at the grid cell level (an area approximately 25 by 25-km in size), which is the native resolution for all climate-related data except for floods and sea level rise, which are evaluated at ~90by 90m resolutions. In this analysis, two scores, namely scores for floods and extreme rainfall and scores for hurricanes and typhoons are used.

The Four Twenty Seven scores are at the address level. Since the mortgage data only contains the first 3 digits of ZIP codes, average 427 scores are calculated for all ZIP codes (first three digits) and used in the regression model along with the severity of hurricanes and floods:

$$PD = \frac{e^{xb+p\ [\ \alpha\ S+\ \beta H+\ \gamma F]*\ C}}{1 +\ e^{xb+p\ \ [\ \alpha\ S+\ \beta H+\ \gamma F]*\ C}}$$

where
p is the probability of flood or hurricane,
S is the average severity of the flood or hurricane,

H is the hurricane score,
F is the flood score,
C is the cure rate factor, defined as 1 minus the probability that the loan will cure,

Monthly probability of flood and hurricanes and the average severity (the mean of severity distribution) are estimated at state level. Since the chance of a loan getting cured after hurricanes and floods is higher than other defaults, a factor is added to account for higher cure rate.

5. Empirical results

To show the impact of hurricanes and floods on mortgage portfolios, three locations with different 427 scores in Texas are considered. For areas with 0 scores the PD increases by a multiplier is 1.01. This means the PDs should be slightly higher due to the higher chance of hurricanes in Texas relative to other states. If the scores are increased to 50 the multiplier will be increased to 1.04 and when the scores are at the maximum value of 100, the multiplier is increases to 1.07 (Table 3). These multipliers could be used in a scenario-based analysis of mortgage portfolios.

Table 3. The impact of 427 scores on mortgage PDs

Hurricane score	Flood score	PD Multiplier
0	0	1.01
50	50	1.04
100	100	1.08

6. Conclusions

Natural disasters have had a significant impact on US mortgage losses. In this paper, we quantified the impact using historical datasets and 427 scores. One of the key elements of incorporating natural disasters in the forecasts is predicting the events at the state level. Due to the relatively small number of events of interest, parametric forms from extreme value distributions have been used. The results show that there is a significant impact on losses when natural disasters are taken into consideration.