Master's thesis - Economics

Ecole Doctorale Sciences Po

Reading about Flood Risk in the News

Evidence from the Housing Market

Jeanne Sorin Le Guevel

Under the supervision of Julia Cagé and Florian Oswald

22^{nd} May 2019

Abstract

Do homeowners know about flood risk? When does this knowledge translate into beliefs, as measured by flood risk discounts on real estate prices? To answer these questions, I built a new dataset of articles from local US newspapers, from which I derive an index of flood risk-related local media pressure in North and South Carolina. This paper proposes a two-step empirical strategy to disentangle the impact of flood risk information shocks on risk awareness and on beliefs. In a first step, I study the impact of local flood-related information shocks on flood risk awareness and, in a second step, on real estate prices. I recover the geographical and temporal heterogeneity awareness using Google Trends. I show that upward flood risk-reclassification induces a discount in real estate prices (1) when properties are valued above $250,000, which is the maximum deductible under the National Flood Insurance Program, and (2) in high awareness regions. Besides, I show that awareness increases when flood risk-related media pressure is substantial. The more dramatic the insurance-related articles published around flood map updates, the larger the impact on both awareness and real estate prices.

Acknowledgments
I am immensely grateful to my supervisors Julia Cagé and Florian Oswald for their time, enthusiasm and invaluable advice. I also want to thank Matthew Kahn and Amine Ouazad, who first introduced me to Geographical Information System analysis and encouraged me to think deeper about environmental risk. Last, I am thankful to Edgard Dewitte, Helene Maghin and Thomas Pellet for their insightful comments and our stimulating discussions, as well as to my family for their unconditional love and support. All remaining errors are mine.

* Contact: jeanne.sorin@sciencespo.fr

Contents

1 Introduction

Real estate assets are the main source of savings for most Americans, but up to $900 billion dollars of real estate investments should be exposed to Sea Level Rise (SLR) risk by 2100[1]. This is equivalent to no less than 13% of household real estate equity losses between 2005 and 2008[2]. While people love to live on the coast and are willing to pay for it, under-pricing current and future, climate change driven increasing flood risk, could lead the coastal property bubble to burst.

I believe the absence of consensus about flood risk capitalization by the housing market in the literature can find its roots in the difficulty of measuring flood risk belief updates through real estate prices. I identify 3 main challenges of this revealed preferences approach. First is the technical difficulty of disentangling positive local amenities and flood risk, as both are likely to be captured by a property's flood risk status. The second challenge comes from the necessity to account for the existence of flood insurance. Its effect could go both ways. On the one hand, the price of a property may reflect the capitalized value of flood insurance premiums (Bin et al. (2008)). On the other hand, flood insurance may create "*a costly and dangerous system of socialized risk to indulge beach lovers*"[3]. A third challenge is to distinguish between flood risk awareness and flood risk beliefs.

In this paper, I contribute to the literature by discussing the importance of this distinction. While the implications are important both for policy and for modeling risk updating in structural models, no paper, to my knowledge, has explicitly studied this difference empirically.

I build a new dataset of flood risk-related articles from 187 local newspapers between 2006 and 2016 in North and South Carolina, two coastal American states increasingly vulnerable to flood risk. I identify 63,170 unique articles related to flood risk and use text mining techniques to automate the analysis of their content. This leads me to consolidate a monthly newspaper panel returning an index of local flood-risk related media pressure, which I combine with a monthly index of flood risk awareness at the Designated Market Level (DMA) using Google Trends. I recover temporal and geographic heterogeneity in flood risk awareness and related media pressure.

This unique database allows me to design a two step identification strategy to (i) identify the sources of flood risk awareness and (ii) study the causal relationship between awareness and beliefs through changes in real estate prices.

In a first step, I delve into the causal relationship between flood risk information shocks and flood risk awareness, as measured by a monthly Google Index. In addition to flood events and flood map updates, I elaborate on Bernstein et al. (2017)'s case study of the 2013 Intergovernmental Panel for Climate Change release and consider a broader definition of information shocks by exploiting the

[1] https://www.zillow.com/research/climate-change-underwater-homes-2-16928/.

[2] https://furmancenter.org/files/publications/HousingandtheGreatRecession.pdf.

[3] https://www.reuters.com/investigates/special-report/waters-edge-the-crisis-of-rising-sea-levels/.

aforementioned local newspaper panel.
In a second step, I move to the capitalization of flood risk into real estate market values. I obtain transaction data from the Zillow Transaction and Assessment Database (ZTRAX)[4]. I extract from this newly available housing database all real estate transactions in the region covered by my awareness index between 2006 and 2016, as well as all houses and owners' characteristics. Using Geographic Information System techniques, I recover all transacted properties' estimated flood risk, as defined by the Federal Emergency Management Agency. I estimate the price of risk by successively running a hedonic analysis and a property fixed effects model. I finally analyze the evolution of flood risk capitalization into the housing market (the evolution of flood risk beliefs) with regards to geographical and time variations in flood risk awareness and related media pressure.

Studying separately awareness and flood risk beliefs allows me to answer two questions.

The first question is *When are people aware of flood risk?* Non surprisingly, floods drive awareness up by 12.0 to 16.1 on Google Index's 0-100 scale. Although I do not estimate such a large impact for any other information shock, I show that floods are not the entire story of flood risk awareness. I underscore the role of flood risk-related local media pressure. When accounting for local newspaper circulation, I find that additional flood related articles, both by themselves and when covering a disaster, increase Google searches, even after controlling for the official disaster damage estimates. Using variations in local newspapers' focus on flood risk-related topics could therefore capture a broader range of "*information shocks*" than the ones traditionally covered by the literature. Regarding the impact of flood map updates on awareness, I find that it is larger in counties covered by more newspapers and not precisely identified in all my specifications. The impact of these updates on Google searches is magnified when accompanied by a large number of insurance related articles with a low sentiment score. I conclude that local newspaper networks are key to spread local official information and that their absence may induce measurable ill-informativeness about environmental risk.

The second question is *When do real estate market participants capitalize flood risk if they know about it?* The underlying reason for formulating the question this way is that homeowners seem to be more aware of flood risk than to believe they are actually at risk themselves.
Although sometimes negative, none of my estimates for the impact of a property's floodplain status on its transaction price is statistically significant when estimated on my full sample. However, I estimate that upward risk-reclassification lead properties valued above $250,000 to appreciate 4.4 percentage points less than non-affected properties. Insights from behavioral economics are crucial to interpret these results. Handel and Schwartzstein (2018) review situations where people fail to use available information to make decisions and insist on the existence of "mental gaps". Limited financial, scientific literacy or a high discount rate could lead people to under estimate flood risk. This holds both for sea level risk risks in the long term (excessive discounts of losses in the far

[4] I am very thankful to Florian Oswald for sharing with me the access to this database.

future) and for annual flood risk in the near future (excessive underestimation of scientific annual flood predictions), which is the focus of this paper's analysis. Because this ability is likely to be correlated with homeowners socio-economic status and wealth, a first part of the answer may be that wealthier homeowners better internalize flood risk. This is consistent with Bernstein et al. (2017), who find that sophisticated investors price sea level risk, but non-sophisticated investors only do so in counties very worried about climate change.

I argue that this is the second part of the answer. I show that properties facing upward risk-reclassification experience a price increase 13.2 percentage points lower than properties remaining low-risk in Designated Market Areas (DMA) with high awareness levels, as recovered from Google Trends. This coefficient is estimated on the full sample of properties sold multiple time, including properties sold for less than $250,000. This penalty for high awareness DMAs drops to 10.3 percentage points when excluding counties that have been flooded in the past, suggesting that flood experience is a key component to accurate flood risk pricing.

Last but not least, disentangling between flood risk awareness and beliefs supports the point made above for the role of insurance. Indeed, among the broad range of information shocks I find having an impact on flood risk awareness, only a handful of them translate into changes in real estate prices and those shocks are related to insurance concerns. Their impact vanishes very fast. Such finding is also consistent with a Bayesian learning model with forgetting (Gallagher (2014)) and with the existence of an availability bias as theorized by Tversky and Kahneman (1974).

To summarize, my findings suggest that local media pressure improves risk awareness, but experience or strong insurance-related concerns are necessary for homeowners to appropriate this abstract knowledge and believe they are themselves at risk.

This paper's contribution to the literature is therefore threefold.

First, it adds to the canonical literature on the real estate price of risk. This literature features ambiguous evidence on the impact of flood risk signals on real estate prices, and especially of floods and flood map status. Indaco et al. (2018) demonstrate that properties in the floodplain sell at a discount in New York but they do not in Miami Beach and Virginia Beach. They even estimate a positive coefficient for the impact of a property's floodplain status on its price in Virginia Beach. In a similar manner, Kousky (2010) does not observe any impact of flood zone status on house prices in St Louis County, Missouri, after the 1993 flood. Other papers attempt to minimize post-flood changes in amenities to focus on changes in flood expectations by looking at neighboring floods not directly affecting a property. These events are interpreted as pure information shocks about a property's underlying flood risk. Hallstrom and Smith (2005) find a price decline of no less than 19% after the 1992 hurricane for properties in flood prone areas that have not been inundated. Bin and Landry (2013) reach similar conclusions in a study on Pitt County, North Carolina.

Second, in line with recent developments of the literature, this paper discusses the relevance of a broader range of flood risk-related information shocks and the role of insurance. Also using the

ZTRAX database but looking at Sea Level Rise, a long-term flood risk, Bernstein et al. (2017) show that the release of the 2013 IPCC report, reevaluating upwards SLR predictions for most US coastal regions, impacted housing prices negatively on the segment of the housing market occupied by the most sophisticated participants. Gibson et al. (2017) also emphasize the importance of information signals more generally as well. The dominant position of the literature is that the flood insurance take-up is too low (Kunreuther (2006), Kunreuther et al. (2009), Gallagher (2014)). However, estimating how much of flood risk capitalization (or absence of capitalization) can be explained by the existence of flood insurance is complex. On the one hand, in a study on Carteret County (NC), Bin et al. (2008) find that most of price discounts associated with the location of a property in the floodplain is captured by the capitalized value of flood insurance premiums. On the other hand, Gibson et al. (2017) study the impact of the Biggert-Waters Act of 2012, which increased flood insurance premia by rulling out a large share of existing premium subsidies. They find that this price signal of risk decreased sale prices of affected properties by about 1.7%, but the estimate was imprecise and much lower than the estimated impacts of Hurricane Sandy in 2012 and of the release of new FEMA floodplain maps.

Third, this paper contributes to the media economics literature. While the role of media coverage and especially newspapers on political outcomes and behavior has been widely covered by Gerber et al. (2009), Gentzkow et al. (2011), Snyder and Stromberg (2010), the literature on the role of media as vectors of environmental risk information transmission is more limited. McCluskey and Rausser (2001) randomly sample a community's main newspaper and show that increased media coverage of a local waste site increased perceived risk and lowered prices of neighboring properties. Freybote and Fruits (2015) focus on the development of underground natural gas transmission pipeline and find that the larger the media coverage of locally unrelated explosions, the more negative the impact on prices of properties further away from the pipeline (with no actual risk). More directly related, Gallagher (2014) exploits the geographic media TV coverage of floods and shows that flood insurance take-up increases in counties located in the same TV market than a flooded county, even when the counties have very different flood histories. My paper distinguishes itself from these studies by exploiting both geographical and temporal variations in media pressure, and by explicitly distinguishing not only media pressure and risk awareness, but also risk awareness and risk beliefs.

The rest of the paper is organized as follows. Section 2 describes the different datasets, Section 3 and 4 present the first and second steps of the identification strategy respectively, and finally Section 5 discusses the results and details possible extensions.

2 Data

In order to estimate the real estate price of flood risk in light of the distinction between flood risk awareness and beliefs, I map real estate transaction data (2.1) to flood risk (2.2) at the property level using Geographical Information System techniques. This consolidated database is leveraged with counties flood history, as recovered from the Presidential Disaster Declarations (2.3), a monthly-DMA awareness index (2.4) and a unique panel of flood risk related news articles at the local newspaper level (2.5) for the 2006-2016 period[5].

2.1 Housing data

I obtain real estate transaction and housing characteristic data from the recently released Zillow Transaction and Assessment Dataset (ZTRAX).

This database is known to be one of the most comprehensive listings of real estate transactions and contains information on many characteristics of a property, including, among others, square footage, rooms, number of bedrooms, number of bathrooms, age of the property, presence of a driveway or a fence.

Table 1: Variables and Summary Statistics

Variable	Description	Mean	St. Dev.	Min	Max	Observations
Structural						
Sales Price	Transaction price of property	243,262	289,582	50,000	10,000,000	621,787
Price per square foot	Transaction price per square foot	247.790	514.055	0.008	61,728.390	621,787
Size	Building size, square feet	2,308	32,611	95	6,642,498	621,787
Age	Building Age	29.986	17.810	7.000	282.000	294,123
Bedrooms	Number of Bedrooms	3.186	0.614	1.000	4.000	435,280
Baths	Number of Bathrooms	2.128	0.537	1.000	3.000	531,931
Driveway	1 = Property with a driveway	0.011	0.104	0	1	621,787
Fence	1 = Property with a fence	0.019	0.137	0	1	621,787
Location						
Distance to Coast	Blockgroup's distance from the coast	89.302	55.816	0.026	205.892	585,617
Floodplain	1 = In an active floodplain	0.036	0.187	0	1	621,787
Block Floodplain	1 = Block partly in the floodplain	0.267	0.442	0	1	621,787
Coast 0.25m	1 = Within 0.25 mile from the coast	0.004	0.065	0	1	585,617
Coast 0.5m	1 = Within 0.5 mile from the coast	0.012	0.109	0	1	585,617
Coast 1m	1 = Within 1 mile from the coast	0.034	0.181	0	1	585,617
Coast 2m	1 = Within 2 miles from the coast	0.074	0.262	0	1	585,617
Coast 5m	1 = Within 5 miles from the coast	0.158	0.365	0	1	585,617

This table includes summary statistics from ZTRAX. Observations are identified at the property-transaction level. Floodplain status is obtain from FEMA flood maps.

In line with Gibson et al. (2017), I filter the ZTRAX database so as to retain only residential

[5] This period corresponds to the common support of my different datasets.

properties[6] between \$50,000 and \$10,000,000. Transactions outside this range are likely to include non-commercial sales, like family transfers (for less than \$50,000 sales) and non-individual sales (for sales above \$10,000,000), which are outside the scope of this paper. In addition, I exclude properties with non-valid geo-coded location or within a Census Blockgroup not mapped on the flood maps[7].

My final sample contains 621,787 transactions (475,183 properties)over 12 DMAs between January 2006 and December 2016. Property structural and location characteristics are described in Table 2.

2.2 Flood risk data

Housing data is combined with flood risk data at the property level by geographically merging the ZTRAX database with the Flood Insurance Rate Maps (FIRM). The FIRMs are produced by the Federal Management Agency (FEMA) in the context of the National Flood Insurance Program (NFIP) enacted by Congress in 1968. As such, it these maps are an established source of flood risk information.

In the 1980s, FEMA released the first hard-copy FIRMs maps for communities participating into the Flood Insurance Program. These hard-copy maps were later digitalized by FEMA and commonly referred to as Q3 maps, or simple Q3. Starting in the beginning of 2000s, these maps have been updated in an effort to improve the accuracy of the estimated flood risk. Because updating these maps is a costly and time intensive process, updates have been on a regular but local basis and some communities still rely on the original Q3 for flood risk prevention. This digital maps are known as the National Flood Hazard Layer, or NFHL. Because it was produced directly as a digital products, the NFHL is considered to be more accurate than the Q3.

The program is also the nation's main flood insurer as more than one million policies insured \$1.2 trilllion of building and contents in January 2017 (FEMA (2017) and Moore (2017)). This corresponds to roughly only half of the properties in the floodplain (Harrison et al. (2001)) and faces a \$250,000 policy per property upper bound. Not all flood risk is therefore covered by the National Flood Insurance Program.

In order to tract the evolution of properties' flood status over time, three digital flood risk data sources were combined for this analysis. I complement my main source for pre-update flood map data, the Q3[8], with an extract of the NFHL as of November 2009 when the Q3 data was missing. I therefore have to exclude from my analysis pre-2009 transactions in communities for which I do not

[6]See Appendix B for a comprehensive filtering rule.

[7]I exclude all Census Tract for which I do not have information on the original flood map status, as on the Q3 shapefile.

[8]I am very thankful to Crowell et al. (2010) for sharing this data with me.

have access to the Q3, as I cannot recover properties' flood status. Third, I merge this pre-update map with the NFHL most recent version, which I extracted from FEMA's website in April 2019[9]. Because FEMA systematically makes most discarded maps unavailable[10], I only have access to the flood insurance map at the three points in time described above.

I make the assumption that maps in each community were updated only once throughout the whole time period. I then map pre-update (Q3 and 2009 NFHL) and post-update (2019 NFHL) properties' flood status to real estate transactions. This assumption relies on a rich literature documenting the very low frequence of flood map updates by FEMA, mainly because of financial constraints (Pralle (2017)). I recover the list of each community's most recent update from the Letters of Final Determination[11] (LFD), released by FEMA, and listing the communities affected by each update[12].

Using Geographic Mapping Software I extract the precise location of the Special Flood Hazard Areas (SFHA) and intersect these map with the geographical coordinates of all properties in my sample. These areas are defined by FEMA as the areas *"that will be inundated by the flood event having a 1-percent chance of being equaled or exceeded in any given year"*. The 1-percent annual chance flood is also referred to as the base flood, or 100-year flood[13]. Figure 1 maps the SFHA coverage for the Greenville-New Bern-Washington DMA (as of April 2019) and illustrates the heterogeneity of the floodplain at a very disaggregated level.

According to the Special Flood Hazard Area definition, buildings falling into this floodplain should theoretically be inundated in the case of a flood similar in intensity to one happening every 100 years according to scientific models.

This statistic should however be taken cautiously for two main reasons. First, this probability should be renewed every year: the probability of not being inundated in a 30 years period is therefore equal to only $0.99^{30} = 0.74$. Second, the National Flood Hazard Layer has been criticized for not accounting for predictions of increasing risks in the future because of both sea level rise and the rise in disasters frequency because of climate change (Pralle (2017)). It should therefore be considered as a lower bound. Flood maps also map a 500-year floodplain, which I exclude from the floodplain variable because flood insurance is never mandatory for houses located in these areas.

[9]`https://data.femadata.com/FIMA/Risk_MAP/NFHL/`.

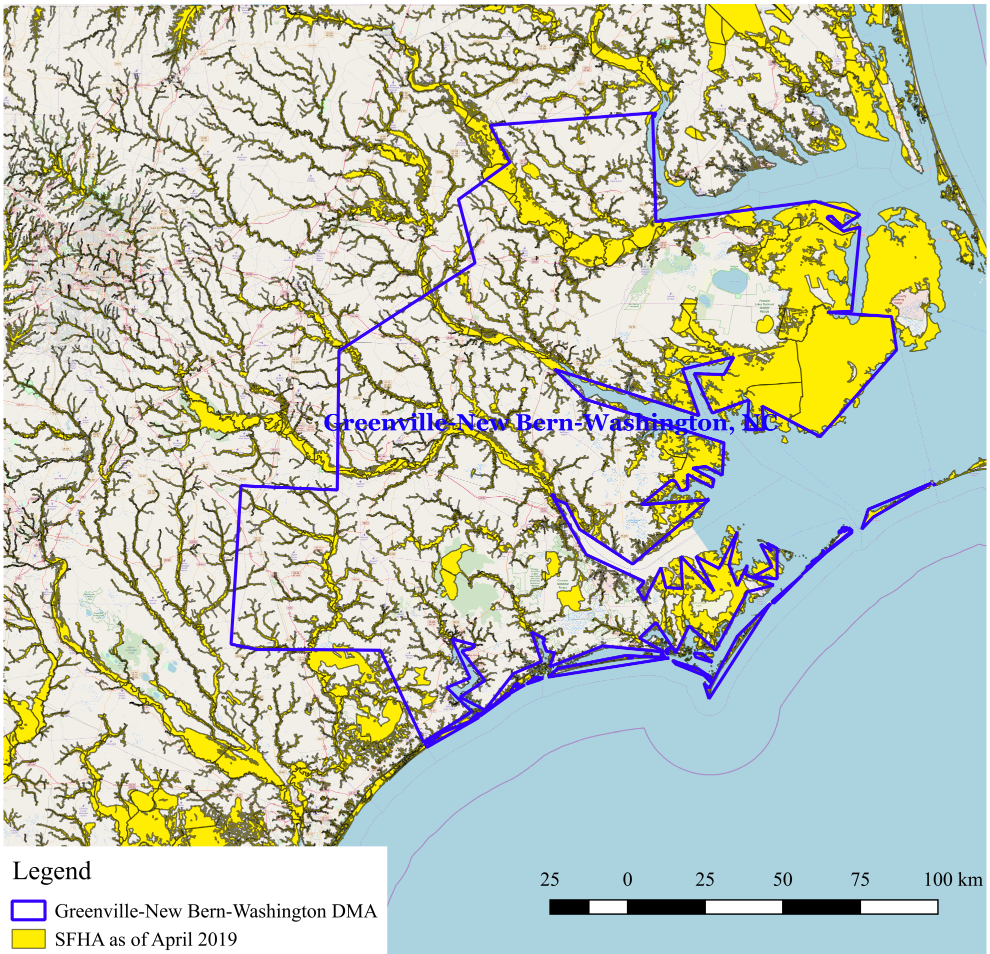
[10]As of May 2019, are only available on FEMA's website the 2019-04-01, the 2018-11-01 and the 2012-09-28 versions. Except the most recent release, no map is complete. Source : `https://data.femadata.com/FIMA/Risk_MAP/NFHL/`.

[11]`https://www.fema.gov/national-flood-insurance-program-community-status-book`.

[12]See Appendix A.3.

[13]Source: `https://www.fema.gov/flood-zones`.

Figure 1: SFHA as of April 2019



This map illustrates the coverage of the Special Flood Hazard Areas (1 percent flood risk areas, SFHA) for the DMA Greenville-New Bern-Washington as of April 2019. Floodplain data are recovered from FEMA.

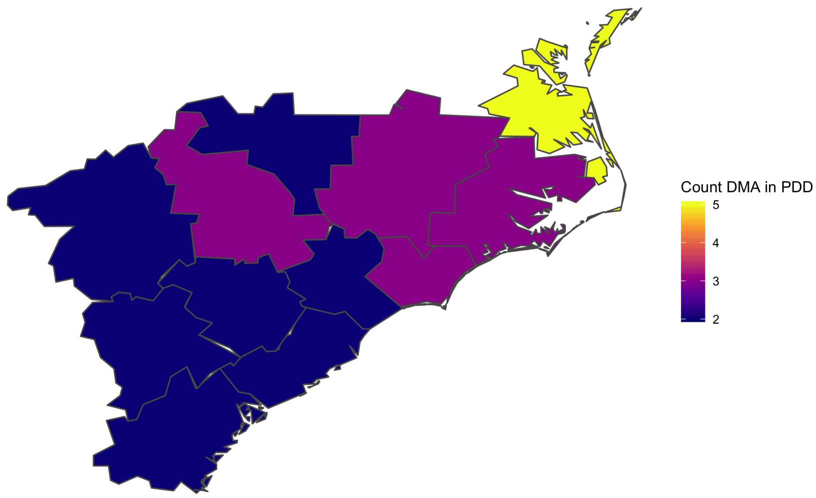
2.3 Flood History

Flood risk history might influence how the housing market capitalizes flood risk in two main ways. First, floods and other natural disasters impact amenities. Second, it has been argued that floods act as flood risk information shocks (Kousky (2010)) .

I recover each county's flood history from FEMA's Presidential Disaster Declarations (PDD) list[14]. In addition to the dates of the disaster, this database contains damage estimates for most PDD, as well as whether each county benefited from Public and/or Private assistance. Figure 2 shows the geographical distribution of these disasters, at the Designated Media Market Level. The distribution of disasters over time is available in Appendix A.4.

[14] https://www.fema.gov/disasters/year.

Figure 2: Number of mentions in PDDs (at the DMA-month level)
2006-2016



This map shows how many months the DMA was mentioned in a PDD for a *Flood* or a *Hurricane* between January 2006 and December 2016. The Norfolk-Portsmouth-Newport News DMA (North-East corner) was in 5 different Flood or Hurricane PDD. PDD data are extracted from FEMA's website.

From FEMA's list of PDDs, I keep the Major Disaster Declarations (DR) but not the Emergence Declarations (EM) because all the disasters in my dataset resulting in an EM also resulted in a DR. Each observation refers to a month-county-disaster pair. This pair exists if a given county was under a PDD for at least one day. For example, a disaster starting on September, $26^{th}$and ending on October 2^{nd} will appear both in September and October. I keep PDD for *Floods*, *Hurricane* because of the flooding and potential flooding respectively associated with these events.

Table 2 presents descriptive statistics of the Presidential Disaster Declaration Data. The average public and individual assistance values are taken from the final declaration. If such information is not available, preliminary estimates from the *Preliminary Damage Assessment Report* [15] are used instead.

The number of houses affected is provided by the Preliminary Damage Assessment report when available. It is otherwise approximated by the number of approved individual applications.

For some disaster declarations, the countywide per capita impact is directly provided by the Preliminary Damage Assessment report. I also compute the average assistance per capita using aggregate final county public and individual assistance variables and county population.

[15] *"The Preliminary Damage Assessment (PDA) process is a mechanism used to determine the impact and magnitude of damage and resulting needs of individuals, businesses, public sector, and community as a whole. Information collected is used by the State as a basis for the Governors request for a major disaster or emergency declaration, and by the President in determining a response to the Governors request"* (44 CFR 206.33).

Table 2: Flood and Hurricane PDD Descriptive Statistics

Statistic	Description	Mean	St. Dev.	Min	Max	N
Public Assistance	1 = county received public assist	1.0	0.1	0	1	243
Individual Assistance	1 = county received individual assist	0.7	0.5	0	1	243
Average Public Assist.	Avg public assist / PDD	3,793,870	2,514,563	0	14,023,770	243
Average Individual Assist.	Avg individual assist / PDD	1,225,362	1,102,480	0	3,757,233	243
Houses Affected	Houses partially destroyed	369.4	254.6	0.0	782.9	243
Impact / Capita	Damage per capita	31.6	40.2	0.1	185.7	49
Individual Assist. / Capita	Individual assist per capita	39.5	64.1	0.0	539.1	243
Public Assist. / Capita	Public assist per capita	126.2	176.3	0.0	1,433.3	243
Flood	1= the PDD is for a "Flood"	0.2	0.4	0	1	243
Hurricane	1= the PDD is for a "Hurricane"	0.8	0.4	0	1	243

This table presents descriptive statistics of the Presidential Disaster Declarations in my sample. PDD are extracted from FEMA's website. PDD are issued a the disaster level and split by county in my dataset. Each observation corresponds to a different county mentioned in a PDD. When available, county specific data is used. PDD level data is used otherwise. All Flood and Hurricane PDD for the 199 counties covered by the Designated Market Areas (DMA) spanning over North Carolina and South Carolina (see Appendix A.1) are included for the 2006-2016 period.

2.4 Google Search Index

In order to measure local flood risk awareness and its evolution over time, I adopt a revealed preferences approach by using Google's measure of the relative popularity of Google searches: Google Search Index (GSI).

This tool has been widely used as a proxy for contemporary awareness in academic research (see for example Choi and Varian (2012), Land and Ryder (2016)), including to measure the evolution of climate change interest in response to changes in temperatures or unemployment (Kahn and Kotchen (2010)).

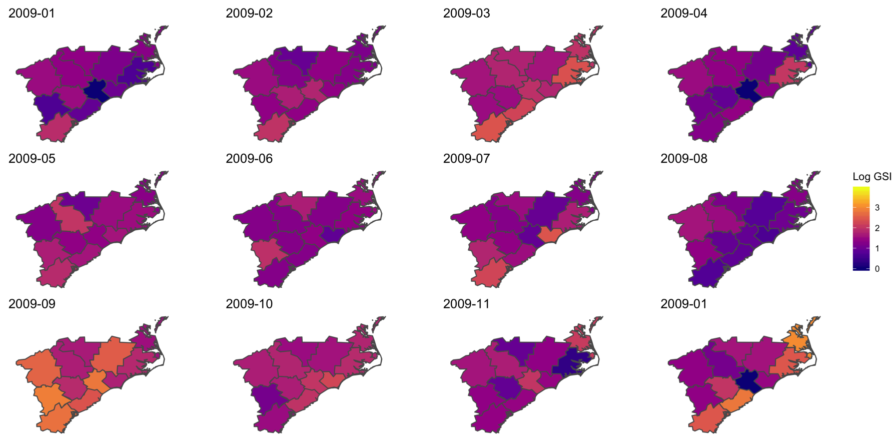
Google Search data presents many advantage for social science and economics researchers. First, it is robust to survey bias, as searchers online are forthcoming and benevolent (Stephens-Davidowitz, 2014). Second, it is a source to track changes in awareness over time and across space.

So far, GSI has mostly been used in the literature on flood risk beliefs for descriptive evidence. Gibson et al. (2017) for example look at state level GSI for one single state, NY. The only paper to exploit the most disaggregated available geographic level, the Designated Media Market (DMA) at the monthly level is Lang and Ryder (2006). They show that experiencing a hurricane increases people's internet searches related to climate change with a lag of a few months.

From Google Trends[16] I download two relative Google Search Index (GSI) at the Designated Market Area (DMA)-month level: one for searches containing the term "Flood", and one for searches containing any of the following expressions: "*Flood*" or "*Flood Map*" or "*Flood Risk*" or "*Floodplain*" or "*Flood Plain*" or "*Flood Layer*" or "*Flood Insurance*".

[16] https://trends.google.fr/trends/.

Figure 3: Google Search Index at the DMA level - 2009



These maps were obtained from data downloaded at `https://trends.google.fr/trends/`. They display the relative Google search intensities of the expressions containing "Flood" by DMA for the DMAs covering North Carolina and South Carolina. Large variations in flood-related searches can be observed.
Darker purple areas correspond to low flood-related google searches, while orange areas reflect peaks in such searches.
Searches being normalized to 100 over the whole period 2007-2017, using a logarithmic scale allows to enhance variations in at lower levels of Google searches. Most periods and DMA areas feature a search index between 5 and 20, while very high relative search index (between 80 and 100) correspond to rare flood events

While the term "Flood" is included in all the terms used to build the second index, both indexes do not hold the same values. Additional investigation into the Google Trends algorithm would be necessary to understand exactly what drives precisely this discrepancy. In a attempt to both capture broader flood risk concerns than would an index built simply on the term "Flood", and avoid capturing searches too unrelated to flood risk concerns as it could be an issue with the second index, I average over these two indexes to create my main Google Search Index. In the rest of the paper I refer to this index as *Google Index (1)*, or simply *Google Index.*

I build a pooled dataset covering the 12 DMAs spanning over North and South Carolina[17] between January 2006 and December 2016.

Google Trends releases its search index on a 0-100 scale: 100 is the maximum search share over the months and geographies of interest. Note that zero-valued observations mean that there were not enough searches of such query in the given period-geography. For example, the datapoint observed for the Charlotte (NC) DMA in May 2010 with value =*10* means that the share of Google requests containing *"flood"* in Charlotte in this month was 10% of the share of similar Google requests in Columbia in October 2015, the maximum GSI (GSI=100), which corresponds to the month of the North American storm complex which caused historic flooding in North and South Carolina. This index does not say anything about the volume of searches.

[17] See Appendix A.1 for a map of geographical coverage.

My dataset features both time and geographic variations: out of my 1534 monthly observations (12 DMA across 11 years), only one equals 100. All shares of the total volume of searches related to these searches are relative to this maximum search share.
The average monthly Google Index is between 3 and 20, with higher indexes only observed on very specific periods including the October 2015 North American Storm Complex and Hurricane Matthew in 2016. Using 2009 as an example, figure 3 highlights that marginal changes in low GSI are important.

Appendix A.5 provides evidence that GSI for the different DMAs do not systematically differ over the period and that flood related searches do not follow traditional seasonal searches like other neutral terms like *wine*. Results do not change significantly if the Google Index is based on the only term *"Flood"*

2.5 Local Newspaper articles

In order to investigate channels and alternative sources of information shocks, I consolidate a unique panel dataset of flood risk related articles from local newspapers in North Carolina, South Carolina, parts of Virginia and of Georgia.
I collect all articles published in 187 local newspapers between 2006 and 2016 containing flood-related terms[18]. These terms were chosen after reading through a large sample of flood and flood map related news articles and picking up related keywords.

Table 3: Sample of Newsbank database

Newspaper	*Bryan County Now*
Date	*August 22, 2013*
Title	*Program focuses on threats to the Georgia coast*
1st paragraph	*Threats to the Georgia coast, ranging from hurricanes to sea level rise, will be the topic of a Ships of the Sea Museum "Coastal Connections" lecture program on Aug. 22, at 7:30 p.m. in the museum's North Garden. University of Georgia Skidaway Institute of Oceanography professor Clark Alexander will present a lecture offering informative and visual program on the hazards facing Georgia's coastal regions. Drawing on two decades of work in the area, he will discuss coastal hazards*
Dynamic summary	*to the Georgia coast, ranging from hurricanes to sea ... level rise, will be the topic of a Ships of the ... Georgia coast, ranging from hurricanes..*

Using data from usnewsdeserts and Factiva's lists of current local newspapers at the county level, I recover a comprehensive list of local newspapers in each 199 counties covered by my anal-

[18]See appendix A.6 for a comprehensive list of such terms.

ysis at each point in time between January 2006 and December 2016. I account for newspaper consolidations and closures. I am able to get articles from 187 local newspapers, out of the 356 of the aforementioned list. I recover circulation estimates from Factiva, *abcas3.auditedmedia.com*, the *American Newspaper Representatives*, the *South Carolina Press Association*, newspapers' websites and personal online investigation[19] for 172 of these 187 newspapers, as well as for 142 of the 169 newspapers not covered in my panel. Comparing the circulation of newspapers in an out of my sample allows me to confirm that all largest newspapers are included in my panel (see Appendix **??**).

For all articles in my sample, I extract the date, the title, the first paragraph, as well as a "dynamic summary", which contains samples of the article featuring the keywords mentioned above. An example of this content is displayed below in Table 3.

For some newspapers, I also know the number of pages and columns of the publication. Unfortunately, my database contains neither the number of words of each article, nor the number of articles published daily/weekly by each newspaper. I am thus neither able to normalize my data by the number of articles published in each newspaper release, nor to weight it by each article's number of words.

Instead, I scale the number of articles published monthly by each newspaper (alternatively in each county) in a similar way as my Google awareness index. This scale goes from 0 to 100, with 100 the maximum number of articles published by a newspaper in a given month.

[19] In this order.

Table 4: Summary Statistics - Newspaper Panel

Variable	Description	Mean	St. Dev.	Min	Max	Count
Newspapers						
Oldest Record	First release on Newsbank	2005	5.617	1985	2016	24,684
Most Recent Record	Last release on Newsbank	2017	2.559	2006	2019	24,684
Circulation	Newspaper circulation estimate	15,381	21,846	400	142,097	22,704
Articles						
Articles	Number of Articles	1.627	4.419	0	136	24,684
Flood Articles	Number of flood articles	0.852	3.057	0	122	24,684
Flood Recovery Articles	Number of flood recovery articles	0.374	1.744	0	77	24,684
Insurance Articles	Number of insurance articles	0.411	1.245	0	37	24,684
Law Articles	Number of law articles	0.280	0.927	0	17	24,684
FEMA Articles	Number of FEMA articles	0.442	1.534	0	42	24,684
Change Articles	Number of change articles	0.624	1.745	0	39	24,684
Flood Map Articles	Number of flood map articles	0.500	1.373	0	28	24,684
Meeting Articles	Number of meeting articles	0.162	0.577	0	12	24,684
Climate Change Articles	Number of climate change articles	0.133	0.676	0	31	24,684
Sentiment Score						
Score	Average monthly sentiment score	-0.543	8.869	-337	96	24,684
Flood Recovery Score	Average score for flood recovery articles	-0.828	5.588	-278	31	24,684
Insurance Score	Average score for insurance article	-0.070	3.638	-108	50	24,684

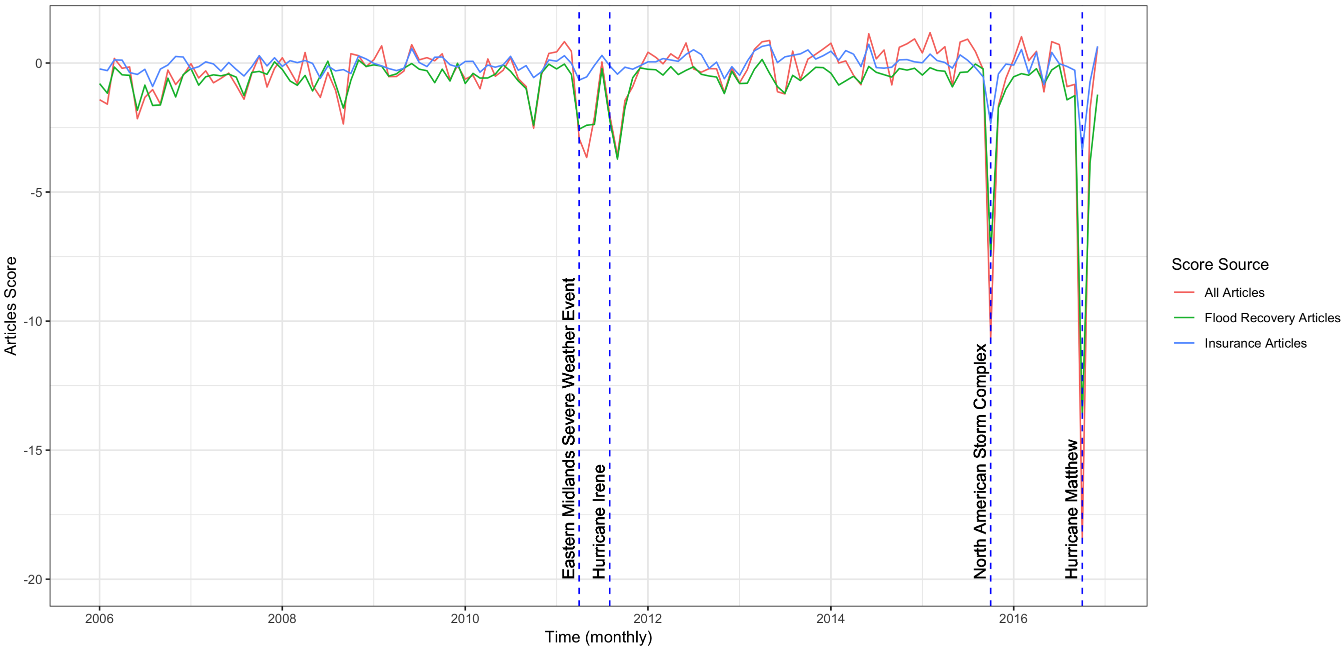
Articles are collected using Newsbank. Circulation data are obtained from Newsbank, Factiva, the South Carolina Press Association and the University of North Carolina. The newspaper panel dataset is at the Newspaper-monthly level and covers 187 newspapers between January 2006 and December 2016.

I analyze the articles' abstracts (title, first summary) using Natural Language Processing (NLP) techniques. I first define 10 categories related to flood risk and classify the articles into these categories according to a word matching algorithm. These categories are *Insurance, Law, FEMA, Change, Flood Map, Meeting, Flood Recovery, Flood, Climate Change* and *Housing*[20]. For example, an article containing at least one of the words *Premium, Insurance, Rate, Ordinance, Insured, Coverage, Covered*, will fall into the Insurance category. Note that these categories are not exclusive. An article may therefore falls into more than one category. When considered separately, the number of articles published for each categories will also be scaled from 0 to 100.

Finally, I extract each article's sentiment score. I argue that this sentiment score, although imperfect as only based on a subset of each articles, reflects the intensity of the experienced flood or flood related events. Ultimately, such measure could be used to calibrate people's reaction to communication tone.

[20]See Appendix A.6 for the list of words belonging to each category.

Figure 4: Articles' sentiment score by category



Article contents is obtained from NewsBank. This figure plots the average article sentiment score, by month, for all articles published by the newspapers in my panel (in red) and for two categories of articles separately : insurance articles (in blue) and flood recovery articles (in green).

Figure 4 shows the evolution of scores for the articles in my sample and for some article categories more specifically. The curve of scores for flood recovery articles follows the aggregate curve, especially on its drops. This is not the case of the insurance articles' score curve. This should not be surprising, as, conditional on being released, articles about flood recovery are more likely to be negative, especially when related to major flood events.

Because scores of flood recovery articles is likely to reflect a flood's seriousness, unlike scores for flood insurance articles, it should not be considered as an exogenous variation in flood-risk related media pressure.

This descriptive evidence motivates the addition of article sentiment score to analyses on flood risk awareness for insurance related articles only.

3 What drives flood risk awareness ?

The literature has traditionally interpreted the impact of floods and flood map updates on real estate prices as the ability of homeowners to account for scientific information and to accurately foresee the impact of rare events on future outcomes.

This approach however assumes that these events are indeed shocks on real estate participants' risk awareness. No work, to my knowledge, has explicitly attempted to disentangle flood risk awareness and flood beliefs or, in other words, the difference between receiving flood information and believing it is relevant to you. Gallagher (2014) is the work with the closest focus, as he showed that news shocks (TV coverage of a flood in a neighbor county) lead to an increase in flood insurance take up almost as large as in the flooded county.

I test this assumption and discuss more broadly the relevance and scope of flood risk information shocks including flood risk related news pressure in the following section.

3.1 Empirical Strategy

I study the impact of 3 types of shocks on contemporary awareness at the DMA-monthly level using Google Trends. This device gives the relative intensities of different Google searches and has been widely used as a proxy for awareness (see Choi and Varian (2012)).

$$\begin{gathered} Google_{dct} = \alpha + \sigma_1 * Flood + \sigma_2 * Map_{ct} + \sigma_3 * News_{ct} + \\ \theta_1 * Flood * News_{ct} + \theta_2 * Map_{ct} * News_{ct} + \\ \gamma_{dcn} + \epsilon_{dct} \end{gathered} \tag{1}$$

Google Search Index *Google* (from 0 to 100) is regressed on a dummy equal to 1 if part of the DMA d is in a Hurricane of Flood PDD *Flood* in month t, on a dummy for a flood map update *Map* in county c in period t, a measure of news pressure *News* at the county-month as well as possibly DMA, county, newspaper and month fixed effects, depending on the specifications.

As described in Section 2.5, I develop two complementary measures of *media pressure* from flood risk-related articles published in local newspapers. First, I scale the number of articles published in a given month-county from 0 to 100 in a similar way to *Google*. Second, I compute a monthly average of these articles' sentiment score at the county (alternatively newspaper) level. Relying on the assumption that for a given flood or map update, variations in the number of newspaper articles (or score) is as good as random after controlling for location and time fixed effects, the coefficients θ_1 and θ_2 give the causal effect of additional media pressure of *Flood* and *Map* respectively on awareness.

3.2 Reduced form results

Table 5 presents the ordinary least squares estimates corresponding to equation (1) on a panel dataset at the county level and aims at isolating the drivers of flood risk awareness. Table 7 proposes a more granular approach as it displays results obtained from a newspaper panel dataset. All specifications have county, year and month fixed effects. All standard errors are clustered at the county level and newspaper level for Table 5 and 7 respectively. *Google Index* correspond to the built measure of Google Index described in Section 2.4 and goes from 0 to 100.

Table 5: The drivers of awareness
Main

	Dependent variable:			
	Google Index			
	(1)	(2)	(3)	(4)
Flood	16.085*** (1.240)		14.441*** (1.411)	
Ln (Flood damage)		1.080*** (0.081)		0.976*** (0.093)
Map Update	0.269 (0.174)	0.267 (0.174)	0.140 (0.179)	0.138 (0.179)
Articles			0.072 (0.047)	0.069 (0.046)
Flood * Articles			0.239** (0.104)	
Ln (Flood damage) * Articles				0.014** (0.007)
Map Update * Articles			0.185 (0.161)	0.185 (0.161)
County FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Month FE	Y	Y	Y	Y
R2	0.281	0.29	0.288	0.296
Within R2	0.169	0.18	0.178	0.187
Observations	26,268	26,268	26,268	26,268

*p<0.1; **p<0.05; ***p<0.01

This table presents the ordinary least squares estimates corresponding to equation (1). The dataset covers 199 counties between 2006 and 2016 (132 months). The dependent variable is *Google Index*, which is an index of relative shares of Google searches at the month-DMA level from 0 to 100. It is defined in Section 2.4 as *Google Index (1)*. *Flood* is a dummy equal to 1 if a county was mentioned in a Presidential Disaster Declaration (PDD) in month t and *Flood damage* is the corresponding damage estimate from the PDD. *Map Update* is a dummy equal to 1 if part of the county's flood map was updated in month t. *Articles* is the number of articles about flood risk published in the county in month t scaled in a similar way to Google Index (from 0 to 100). All specifications include year, month and county fixed effects. Standard errors are presented next to estimates and are clustered at the county level.

In Table 5, Google Index is regressed on a measure of flooding and a dummy equal 1 if a map update happened in county i in month t.

The positive significant coefficients for *Floods* in all specifications confirm that disasters can be considered as *Information shocks* as they affect people's flood risk awareness, as measured by *Google Index*. The coefficients for map update are positive but not statistically significant at traditional confidence intervals. Moreover, the coefficients for floods and map updates are of different magnitudes. Where a flood causes Google Index to jump by 16.085*** points (on a 0-100 scale) in Designated Media Markets (DMA) where at least one county was mentioned in the Presidential Disaster Declaration, a map update only increases this index by 0.269(0.174) (significant at the 88% confidence interval). In columns (3), and (4), the specifications from (1) and (2) are augmented with covariates for the scaled number of flood risk related articles published in county i in month t.

When the model is run on the monthly panel of newspaper articles at the county level (Table 11, the coefficients for *Articles* and *Map Update * Articles* are not statistically significant. Weighting the number of articles by newspaper circulation suggests that the lack of statistical significance in the previous table may be due to an imprecise specification.

Table 6: The drivers of awareness
Weighted

	Dependent variable:	
	Google Index	
	(1)	(2)
Flood	12.530*** (1.514)	
Ln (Flood damage)		0.843*** (0.100)
Map Update	0.415* (0.234)	0.414* (0.234)
Articles (weighted)	0.117*** (0.023)	0.113*** (0.022)
Flood * Articles (weighted)	0.115 (0.110)	
Ln (Flood damage) * Articles (weighted)		0.007 (0.007)
Map Update * Articles (weighted)	0.207 (0.201)	0.206 (0.201)
County FE	Y	Y
Year FE	Y	Y
Month FE	Y	Y
R2	0.322	0.329
Within R2	0.198	0.206
Observations	12,624	12,624

*p<0.1; **p<0.05; ***p<0.01

This table presents the ordinary least squares estimates corresponding to equation (1). Unlike in Table 5, counties with no newspaper or missing circulation data for newspapers in the panel are excluded from the analysis. The aggregated (weighted) number of articles at the county-month level is then normalized from 0 to 100. Observations are weighted by the share of the county's recovered circulation (circulation of newspapers in the panel / total estimated county circulation). Standard errors are presented next to estimates and are clustered at the county level.

Indeed, when normalizing the number of articles by the estimated newspaper circulation and weighting observations at the county level by the share of the county's recovered circulation (circulation of newspapers in the panel / total estimated county circulation), the coefficient for *Articles* is positive and statistically significant at the 99% confidence interval (0.117***). This coefficient suggests that there exists other flood risk related information shocks than the ones previously considered. One could think of floods too small to lead to a PDD, public discussions about flood risk or newspaper investigations about flood preparedness.

That the coefficient for *Map Update* becomes significant in this specification deserves some comments. Counties with missing circulation of with 0 newspaper are excluded from the panel. Thus, a larger, or statistically significant coefficient for *Map Update* in Table 6 columns (1)-(2) than in Table 5 columns (3)-(4) suggest that the impact of map update on awareness, and the a county's newspaper coverage are correlated.

Following this hint about the importance to account for newspaper coverage (the number of newspapers per county), and in order to improve the accuracy of my analysis, I then run the same model on a newspaper panel dataset (observations at the newspaper-month level). Results are presented in Table 7. Columns (1) and (2) correspond to the specifications presented in Table 5's columns (2) and (4) and columns (3) and (4) consider separately Insurance- and Flood recovery-related articles.

The coefficient for *Map Update* is positive and statistically significant. This is consistent with the intuition developed above that giving more weights to counties with more newspapers (the newspaper panel has one observation per newspaper-month) leads to estimate a positive impact of map updates on flood risk awareness. I will be interested in studying whether an increasingly documented phenomenon like the emergence of "news deserts" has a systematic impact on awareness, these deserts being likely correlated with other local and community degrees of public life.

The interaction between the number of articles and the flood dummy is positive and significant, and robust to the replacement of the flood dummy by an estimate of flood damage.

Finally, in columns (2)-(4), I exploit differences in articles' content and tone. I split the number of articles by articles related to flood recovery and articles related to flood insurance, and add to the latter the article's *Score*, a measure of the article's tone. Both the coefficient for flood recovery article and the interaction between flood recovery articles and flood damage are statistically significant (0.113^{***} and 0.013^{**} respectively). The number of insurance articles and the interaction of map update with the number of insurance articles weighted by their sentiment score have statistically significant impact on Google searches (0.025^{*} and -0.021^{***} respectively). For a given map update in month t, and a given number of flood insurance-related articles published by a given newspaper in this month, a drop of the articles' score from 0 to -10 leads to an increase in flood risk awareness, as measured by Google Index, of 0.21. While this coefficient may seem small at first, it means that if 10 articles were published by newspaper i in month t with an average sentiment score of -10, awareness rises by more than 2, in addition to the map update impact on Google Index itself. No other article category's score interacted with *Map Update* lead to similar estimates (Table 19), suggesting that flood risk awareness is mainly driven by insurance concerns.

Table 7: The drivers of awareness
Newspaper panel dataset

	Dependent variable:			
	Google Index			
	(1)	(2)	(3)	(4)
Flood	11.964*** (1.198)		12.335*** (1.213)	
Ln (Flood damage)		0.809*** (0.078)		0.833*** (0.079)
Map Update	0.527*** (0.190)	0.524*** (0.190)	0.577*** (0.164)	0.574*** (0.164)
Articles	0.093*** (0.012)	0.091*** (0.012)		
Flood *Articles	0.283*** (0.080)			
Ln (Flood damage) * Articles		0.017*** (0.005)		
Map Update * Articles	0.053 (0.144)	0.054 (0.144)		
Flood Recovery Articles			0.116*** (0.026)	0.113*** (0.026)
Insurance Articles			0.025* (0.015)	0.025* (0.015)
Insurance Articles * Articles score			−0.002 (0.002)	−0.002 (0.002)
Flood * Flood Recovery Articles			0.222** (0.104)	
Ln (Flood damage) * Flood Recovery Articles				0.013** (0.007)
Map Update * Insurance Articles			0.007 (0.058)	0.007 (0.058)
Map Update * Insurance Articles * Score			−0.021*** (0.005)	−0.021*** (0.005)
County FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Month FE	Y	Y	Y	Y
R2	0.328	0.333	0.335	0.34
Within R2	0.217	0.223	0.225	0.232
Observations	24,684	24,684	24,684	24,684

*p<0.1; **p<0.05; ***p<0.01

This table presents the results of equation (1) estimated on a monthly newspaper panel of 187 local newspapers between 2006 and 2016. The dependent is *Google Index*, which is an index of relative shares of Google searches at the month-DMA level from 0 to 100. *Flood* is a dummy equal to 1 if a county was mentioned in a Presidential Disaster Declaration (PDD) in month t and *Flood damage* is the corresponding damage estimate from the PDD. *Map Update* is a dummy equal to 1 if part of the county's flood map was updated in month t. *Articles* is the scaled number of articles about flood risk published by newspaper i in month t. This scale goes from 0 to 100, with 100 the maximum number of articles (per category) published by a single newspaper i in a given month t. In columns (3) and (4) I consider separately articles about insurance and about flood recovery and add a measure of insurance article's sentiment score (the lower the more negative the article). All specifications are presented with year, month and county fixed effects. Standard errors are presented next to estimates and are clustered at the Newspaper level.

3.3 Robustness Checks

I check that these conclusions are robust to modifications of the specifications.

More precisely, I run the analysis without county fixed effects (Tables 17 and 18) and results are overall consistent with Tables 5 and 7 respectively. Notably, when one does not control for county fixed effects, the coefficient for *Map Update* is no more significant at conventional confidence intervals when the model is run on the county panel. It is still positive and statistically significant for the newspaper panel model (0.385** in column (3)). Besides, the positive significant coefficient for the interaction between *Flood* and *Flood Recovery Articles* resists to the addition of flood damage measures for the second awareness measure.

I also run the the newspaper panel analysis on a panel where each observation is weighted by the newspaper's circulation. Results are presented in Table 17 and do not systematically differ from the main specification displayed in Table 7.

4 Flood risk and the housing market

People respond to flood risk related information shocks by increasing the relative share of flood risk related Google searches. In this section I investigate whether homeowners price flood risk in their real estate decisions and whether heterogeneity in awareness can explain uneven flood risk capitalization.

4.1 Empirical Strategy

4.1.1 The value of being located in the floodplain

All else equal, a property located in the Special Flood Hazard Area should sell at a discount compared to a property located outside the floodplain. This claim assumes that FEMA designated floodplain, the SFHA, is believed to be a credible prediction of a property's actual flood risk, and that participants in the real estate market discount the associated future potential losses.

The empirical strategy is designed such as to compare properties with equivalent characteristics located in the same neighborhood, and selling in the same time period in order to isolate the causal effect of flood risk status on real estate prices.

Because the Zillow Transaction Assessment Dataset contains a large number of property and building attributes that help to explain a property's price, I am able to run a full hedonic analysis on house prices to estimate the value of being located in the floodplain.

Following the hedonic analysis methodology developed in Rosen (1974), my main specification is:

$$ln(Y_{it}) = \beta_t FP_{it} + X_{it}\phi + \lambda_{FE_{itz}} + \epsilon_{it} \tag{2}$$

The dependent variable $ln(Y_{it})$ is the log of the transaction price for property i in period t, or alternatively the log of the transaction price per square foot. The period is defined at the year-month level so as to match my other datasets' time definition. The explanatory variable of interest,

FP_{it} is a dummy equal to 1 if property i is in a floodplain active at the time of sale t. This encompasses properties in the floodplain on the old map sold before the map update, properties in the floodplain on the new map sold after the update as well as properties in the floodplain on both map versions. X_{it} includes 4th order polynomials of the property's area in square meters and its age[21]. Finally $\lambda_{FE_{itz}}$ is at the core of the identification strategy. It is a matrix of all interacted fixed effects for neighborhood z defined as Census Tract; distance to the coast buffer[22]; property type (single-family or multiple family dwelling, apartment, condominium, mobile home...); number of bedrooms b; number of bathrooms a and whether the aforementioned size measure designates area of the gross full building, of the living or heated space etc[23] The addition of distance to the coast fixed effect to more traditional interactions is guided by Bernstein et al. (2017) identification strategy. They show that the value of properties increases non-linearly with their proximity to the coastline, even after controlling for neighborhood and time fixed effects and explain this relationship by the presence of coastal amenities and local accesses to the beach. Figure 5 illustrates the same relationship for my sample. In order to identify the causal impact of flood risk status, also positively correlated with proximity to the coast, distance to the coast must therefore be interacted with time and location fixed effects. β can then be interpreted as the causal effect of an active floodplain status on house prices if one accepts that the remaining variations in flood risk status is random. My main specification is ran on properties within 0.25 mile from the coast. This distance threshold follows directly from Bernstein et al. (2017) and is chosen to be the optimal geographical limit for the specification with interacted fixed effects including distance-to-coast buffers.

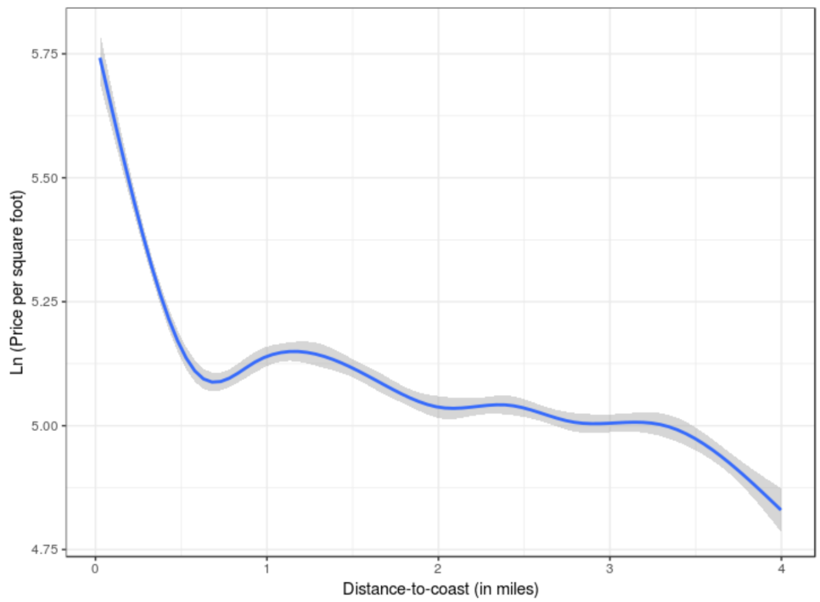
[21] I do not have a precise enough measure of distance to the coast to be able to add it as covariate in addition to coast distance buffers fixed effects.

[22] Cutoffs are every 0.05 mile up to 0.5 mile, every 0.1 mile up to 10 miles.

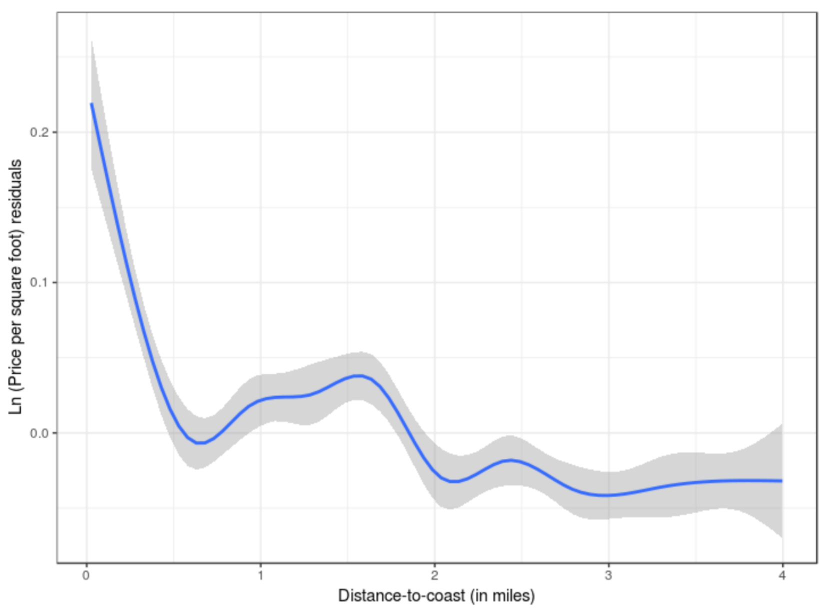
[23] Bernstein et al. (2017) add occupancy status, condominium indication and elevation bucket above sea level. Unfortunately, as of today, I do not have access to these characteristics.

Figure 5: A case for distance-to-coast fixed effects

(a)



(b)



Notes: Figure 5 presents supporting evidence for the choice of a specification with distance-to-coast fixed effects. Figure (a) displays the relationship between the log of the transaction price per square foot and the distance to the coast using local polynomial regression fitting methods. Figure (b) is obtained with similar method and presents the relationship between the residual of the log transaction per square foot on Census Tract and month interacted fixed effects, and the distance-to-coast in miles.

4.1.2 The impact of risk-reclassification on house prices

Whether properties in the floodplain sell at a premium or a discount could still capture differences in nature between these properties and properties not classified at risk if non-observed house characteristics (orientation, soil, water system) were to increase both its hedonic value and flood risk.

To address this potential omitted variable bias, the literature has exploited flood events (Kousky (2010), Kunreuther (2006)).

These findings are consistent with the existence of an *availability bias* first theorized by Tversky and Kahneman (1974) stating that one reason why people do not account for all existing information in forming their beliefs about the likelihood of uncertain events is because this information in not readily available. However, while these events can be interpreted as *information shocks* about a property's underlying flood risk, they also affect amenities by potentially damaging both the property and its neighborhood.

Alternatively, flood map updates carry new information without changes in neighboring amenities (Kousky (2010) , Indaco et al. (2018))[24]. As argued before, these flood map updates are potentially valid information shocks as they induce an increase in flood risk awareness as measured by *Google Index*[25]. This information transmission is especially important when newspapers emphasize the implications with regards to expensive flood insurance.

In order to identify the effect of map updates, which imply changes in flood insurance premium for affected properties, I follow Indaco et al. (2018) and Hallstrom and Smith (2005) and run a unit fixed effect model with a post coefficient on the panel of repeated sales that bracket the map updates. More precisely, I sample my transaction database so as to keep only properties sold at least once before the update and once after (not all communities were updated at the same time, so I have in practice as many treatments as different map updates). If a property was sold more than twice, I keep the closest sale on each side of the update date.

Accordingly, I classify the properties into categories: properties entering the floodplain after the map update (*OutIn*), properties leaving the floodplain after the update (*InOut*), properties always in the floodplain (*InIn*) and properties never on the floodplain (*OutOut*).

Similarly to Indaco et al. (2018), my main specification is a model with property (unit), year and month (time) dummies, as well as a "Post" treatment dummy.

$$ln(Y_{izt}) = \sigma_i + \lambda_t + \theta Post_{izt} + \beta_t Post_{izt} * Category_i + \epsilon_{izt} \tag{3}$$

Unit fixed effects σ_i are defined at the property level and do not interact with year fixed dummies

[24]New flood map may require that communities update their flood resilient installations (levees).

[25]Statistically significant at the 88% confidence interval when estimated on an unweighted county panel at the monthly level. See Table 5.

λ_t. *PostUpdate* is a dummy equal to 1 if the property is sold after its community map update and *Category* is a categorical variable with a property i's four possible map status: *InIn*, *OutOut*, *OutIn*, *InOut*.

This specification assume that property fixed effects are time invariant. It is useful to think about this assumption as the unit fixed-effect version of the canonical difference-in-differences' common trend assumption, but adapted to a *difference-in-differences model with unit fixed effects* by conditioning on on the unit rather than the group.

For the sake of clarity, I run this *difference-in-differences with fixed effect model* on two different samples of the repeated sales panel. First, I compare *OutOut* properties to *OutIn* properties. I expect to find a negative coefficient for the interaction *Post * OutIn* if an upward reevaluation of a property's flood risk (*OutIn*) has a negative impact on its price. Second, I focus on *InIn* and *InOut* properties. I expect to find a positive coefficient for the interaction *Post * InOut* if a downward reevaluation of a property's flood risk (*InOut*) has a positive impact on its price.

4.1.3 The role of awareness

A general approach on awareness and risk reclassification

In Section 3, I have discussed the impact of flood-related information shocks traditionally considered by this literature (floods and map updates) on flood risk awareness and shown that some of these shocks' heterogeneous impact on flood risk awareness were partly driven by media pressure and coverage of these events.

I propose here to exploit this heterogeneity in the effectiveness of information shocks, and more specifically flood map updates, to understand whether awareness and beliefs are related one-to-one.

Few studies have attempted to estimate whether differences in local flood risk awareness may drive flood risk premia/discounts on house prices or, in other words, the impact of awareness on flood risk beliefs as measured by real estate prices.

Using the Yale Climate Opinion data providing county level survey data of flood climate change concern, Bernstein et al. (2017) find that owner-occupied houses in counties with a reported score of worry in the highest decile sell at a 10% discount when exposed to Sea Level Rise risk.

Sea Level Rise risk is of a different nature than the flood risk measured by the Special Flood Hazard Areas (SFHA) as the former is long term (30-50 years) while the latter is short term and continuous. Rather than concerns about climate change, looking at real estate participants' concerns about contemporaneous climate risk is therefore more relevant here.

I use the same unit fixed effect model described above (Equation (3)), which I augment with *High Awareness DMA*, a dummy equal to 1 if the average Google Index in the property's DMA is above the median of Google Index averages and *High Awareness Month*, a dummy equal to 1 if the monthly-DMA Google Index is above the 75^{th} percentile of the sample's monthly-DMA indexes.

Eliciting a candidate mechanism: focus on the flood map updating process

Lastly, I delve into the flood map updating process itself. Updating a flood maps is not a one-time event, but a process starting months before the map actually becomes effective. This timeline is summarized in figure 6 below.

Figure 6: Flood Map Update Timeline

Ten to Twelve months before the new maps finally becomes effective, FEMA releases the Preliminary Base Flood Elevations (Preliminary BFE). Within a two weeks period, it then publishes a notice of this release in two local newspapers. Then starts a ninety days appeal period where homeowners and communities can submit map corrections or appeals. At the end of this period, FEMA will finalize the maps and publish a *Letter of Final Determination*. The new map will be pending during 6 months. By the end of this period, homeowners and communities must have complied with all legal requirements, including getting a flood insurance when buying a house with a government-backed mortgage[26].

In an event study, I regress transaction price on each period of the map updating process and isolate the additional effect of high awareness months.

$$ln(Y_{it}) = \beta_t FP_{it} + X_{it}\phi + \lambda_{FE_{itz}} + \sum_{j=1}^{J} \beta_j Period_{itj} + \sum_{j=1}^{J} \theta_j Period_{itj} HighAwareness_{zt} + \epsilon_{it} \quad (4)$$

Where J indexes the different periods properties experiencing risk reclassification go through: *Pre Update Process* ; *Preliminary Base Flood Elevation* ; *Appeal Period* ; *Map Pending* ; *Map Update Month* and *Post Update.*. The sample is restricted to *OutIn* properties.

[26] `https://www.fema.gov/media-library-data/1468504201672-3c52280b1b1d936e8d23e26f12816017/Flood_Hazard_Mapping_Updates_Overview_Fact_Sheet.pdf`.

4.2 Reduced Form Results

In this section, I first discuss the results of the hedonic analysis before moving to a difference-in-differences set up. Finally, I run an event study on the flood map updating process in an effort to discuss the role of awareness and insurance concerns in the pricing of flood risk on the housing market.

4.2.1 The value of being located in the floodplain

I present the full results of the hedonic model (Equation (2)) in Table 8 and show that houses in the floodplain do not sell at a premium or discount within 0.25 miles from the coastline.

I regress the log of the property's transaction price on different combinations of covariates and fixed effects. The absence of statistically significant coefficient in columns (1)-(4) suggests that floodplain status does not on average have an impact on a property's value on the real estate market. Column (6) presents results for the same specification as in column (2) but the dependent variable is the log of the price per square foot.

Column (5) is included to emphasize the importance of the hedonic specification with interacted fixed effects to isolate the causal relationship between floodplain status and house prices, and more specifically to account for the positive correlation between coastal amenities and flood risk.

Table 8: Full Regression Results - Main

	Dependent variable:					
	Ln Price					Ln Price per square foot
	(1)	(2)	(3)	(4)	(5)	(6)
Floodplain	0.079	0.074	0.114	0.028	0.308***	0.074
	(0.077)	(0.081)	(0.235)	(0.113)	(0.026)	(0.081)
Ln(Square Feet)		53.267**	71.901	58.641		52.267**
		(23.680)	(65.079)	(48.347)		(23.680)
Ln(Square Feet)^2		−10.806**	−14.507	−11.920		−10.806**
		(4.700)	(12.681)	(9.394)		(4.700)
Ln(Square Feet)^3		0.955**	1.277	1.055		0.955**
		(0.407)	(1.080)	(0.799)		(0.407)
Ln(Square Feet)^4		−0.031**	−0.041	−0.034		−0.031**
		(0.013)	(0.034)	(0.025)		(0.013)
Ln(Property Age)			−35.131			
			(69.463)			
Ln(Property Age)^2			18.330			
			(36.243)			
Ln(Property Age)^3			−4.196			
			(8.276)			
Ln(Property Age)^4			0.353			
			(0.698)			
T * L * D * B	Y	Y	Y			Y
T * L * D * B * BF * U				Y		
Location FE	Tract	Tract	Tract	Tract		Tract
R2	0.818	0.822	0.892	0.901	0.06	0.933
Within R2	0.005	0.027	0.112	0.032	0.06	0.769
Observations	2,278	2,278	1,285	2,278	2,291	2,278

*p<0.1; **p<0.05; ***p<0.01

This table presents the ordinary least squares estimates corresponding to equation (2). The explanatory variable of interest is *Floodplain*, which is a dummy equal to 1 if the transacted property is in a (active) floodplain. Real estate transaction data and house characteristics are from Zillow. Floodplain data are from FEMA. I trim the real estate dataset to exclude the bottom and top 1% sales amount. The sample is then restricted to properties located in Census block groups within 0.25 miles from the coastline. This corresponds to 2,291 transaction, among which 1,285 transaction have data for the property's age, and 2,278 have data about the number of bedrooms. In columns (1) to (5) the dependent variable is the log of the transaction price. Columns (1) to (3) and (6) include time (year-month) (T), location (Census Tract) (L), distance to the coast buffer (every 0.05 mile up to 0.5 mile, every 0.1 mile up to 10 miles) (D) and number of bedrooms (B) fixed effects, as well as their interactions. Column (4) also has two additional fixed effects interacted with the ones mentioned above: Block in the floodplain (BF) equal to 1 if at least one property on the block is in the floodplain (even if the property itself is not) and property use type (single family house, multiple family dwelling, condominium...) (U). No covariate or fixed effect is included in column (5). The inconsistent coefficient for *floodplain* emphasizes the importance of our preferred specification with covariates and interacted fixed effects. In column (6) the dependent variable is the log of the transaction price per square foot. The specification is similar to column (2). Standard errors are presented below estimates and are clustered at the location fixed effect level.

Table 9 presents the ordinary least squares' main estimate for different distance-to-coast cutoffs. All specifications are identical to column (2) of Table 8. Positive statistically significant floodplain coefficients in column (2) onward should be taken cautiously. As mentioned above, the fit of the specification is expected to decrease when properties are located further away from the coast, as precise distance to the beach is not a key component of a property's hedonic price anymore. The floodplain status coefficient is therefore likely to contain other time-invariant, local amenities or unobservable characteristics.

Table 9: Full Regression Results - Varying distance-to-coast cutoff

	Dependent variable:					
	Ln Price					
	(1)	(2)	(3)	(4)	(5)	(6)
Floodplain	0.074	0.092*	0.056***	0.068***	0.068***	0.062***
	(0.081)	(0.051)	(0.016)	(0.013)	(0.014)	(0.010)
Size Controls	Y	Y	Y	Y	Y	Y
Age Controls						
T * L * D * B	Y	Y	Y	Y	Y	Y
T * L * D * B * BF * U						
Location FE	Tract	Tract	Tract	Tract	Tract	Tract
Maximum distance to the coast	0.25 mile	0.5 mile	1 mile	2 miles	5 miles	10 miles
R2	0.822	0.872	0.835	0.828	0.831	0.83
Within R2	0.027	0.03	0.051	0.054	0.07	0.068
Observations	2,278	6,515	18,400	39,874	84,772	104,080

*p<0.1; **p<0.05; ***p<0.01

This table presents the ordinary least squares estimate for the explanatory variable of interest, *Floodplain*, corresponding to equation (2) with samples restricted to properties within 0.25, 0.5, 1, 2 and 5 miles from the coast. Real estate transaction data and house characteristics are from Zillow. Floodplain data are from FEMA. I trim the real estate dataset to exclude the bottom and top 1% sales amount. Column (1) is similar to column (2) of Table 8. All columns have the same specification and include 4th order polynomials for building size as covariates as well as simple and interacted fixed effects for time, location, distance to the coast buffer (every 0.05 mile up to 0.5 mile, every 0.1 mile up to 10 miles) and number of bedrooms. Standard errors are presented below estimates and clustered at the location fixed effect level.

In order to address this issue, I estimate the same model but keeping in my sample properties located in blocks containing at least one property in the floodplain only. Under the assumption that within a Census block, being located in the floodplain or no is as good as random, the floodplain coefficient should be estimated more accurately. Results are presented in Table 10.

Table 10: Main Regression Results - Only Blocks in the floodplain

	Dependent variable:					
	Ln Price					
	(1)	(2)	(3)	(4)	(5)	(6)
Floodplain	−0.290	−0.062	−0.035	0.026	−0.011	−0.002
	(0.657)	(0.344)	(0.115)	(0.083)	(0.057)	(0.048)
Size Controls	Y	Y	Y	Y	Y	Y
Age Controls						
T * L * D * B	Y	Y	Y	Y	Y	Y
T * L * D * B * BF * U						
Location FE	Tract	Tract	Tract	Tract	Tract	Tract
Maximum distance to the coast	0.25 mile	0.5 mile	1 mile	2 miles	5 miles	10 miles
R2	0.88	0.911	0.895	0.891	0.899	0.902
Within R2	0.058	0.018	0.029	0.04	0.035	0.034
Observations	1,010	1,909	4,800	8,830	15,282	17,410

*p<0.1; **p<0.05; ***p<0.01

This table presents the ordinary least squares estimate for the explanatory variable of interest, *Floodplain*, corresponding to equation (2) with samples restricted to properties within 0.25, 0.5, 1, 2, 5 and 10 miles from the coast. Real estate transaction data and house characteristics are from Zillow. Floodplain data are from FEMA.I trim the real estate dataset to exclude the bottom and top 1% sales amount. Unlike Table 9, properties located in Census blocks containing at least 1 property in the floodplain only are included in the sample. This robustness check is aimed at addressing refining the comparison between treated and control properties.
All specifications include 4th order polynomials for building size as covariates as well as simple and interacted fixed effects for time, location, distance to the coast buffer (every 0.05 mile up to 0.5 mile, every 0.1 mile up to 10 miles) and number of bedrooms. Standard errors are presented below estimates and clustered at the location fixed effect level.

Unlike estimates presented in Table 9, estimates of the *Floodplain* coefficient for the sample including only properties in *risky* blocks are negative, although not significant ($-0.290(0.657)$) for the 0.25 mile-to-coast cutoff). This suggests that while the main specification may fail appropriately control for time invariant local amenities more than 0.25 mile away from the coast, in *risky* blocks, properties in the floodplain do not sell at a premium, and potentially sell at a discount.

4.2.2 The impact of risk-reclassification on house prices

Estimating a difference-in-differences model with fixed effects on the panel of repeated sales instead should correct for potential caveats of the hedonic analysis brought to light by the last specification. Results are presented in Table 11. My panel of repeated sales contains 34,522 transactions (17,261 properties) sold between 2006 and 2016.

Table 11: Risk Reclassification Impact on House Prices
Main

	Dependent variable:			
	Ln Price			
	(1)	(2)	(3)	(4)
Post Update	0.181***	0.015**	0.145***	−0.005
	(0.007)	(0.007)	(0.053)	(0.073)
Post Update * OutIn	−0.016	−0.045**		
	(0.019)	(0.018)		
Post Update * InOut			0.049	0.103**
			(0.037)	(0.047)
Categories	OutOut, OutIn	OutOut, OutIn	InIn, InOut	InIn, InOut
Minimum Transaction Price		250,000		250,000
Ref cat	OutOut	OutOut	InIn	InIn
Year dummy	Y	Y	Y	Y
Month dummy	Y	Y	Y	Y
Property FE	Y	Y	Y	Y
R2	0.871	0.89	0.889	0.865
Within R2	0.055	0.06	0.073	0.253
Observations	45,002	10,918	944	278

Note: *p<0.1; **p<0.05; ***p<0.01

This table presents the results of the unit fixed effect model summarized in equation (3) estimated on a panel of properties that were sold both before and after their community's map update. Real estate transaction data and house characteristics are from Zillow. Floodplain data are from FEMA. I trim the transaction to exclude the bottom and top 1% sales amount. When a property is sold more than once before or after the update, only the closest sale to the update is considered. The resulting panel dataset covers 45,946 transactions (45,002+944), i.e. 22,973 properties. In column (1) and (2), *OutOut* and *OutIn* are considered. The reference category is the *OutOut* properties sold before the update. In column (3) and (4), *InIn* properties are compared to *InOut* properties. All specifications include property, fixed effects as well as year and month dummies. Unlike odd columns, results presented in even columns are estimated on a subset of the sample excluding properties sold for less than $250,000. Standard errors are presented below estimates and are not clustered.

For the sake of clarity I split my analysis into two parts. I compare *OutOut* and *OutIn* properties in the first two columns to estimate the impact of an upward risk-reclassification on real estate prices. *OutOut* is the reference category. In the last two columns I compare *InIn* and *InOut* properties to estimate the impact of a downward risk-reclassification on real estate prices. *InIn* is the reference category. All specifications include property, year and month fixed effects.

Column (1) reveals that *OutOut* properties appreciated on average by 20% ($exp(0.181)$) between the first and second sale. The coefficient for the interaction between *Post Update* and *InOut* in column (1), -0.016 is not statistically significant. However, this coefficient falls to $-0.045(0.018)$ and becomes statistically significant at the 95% confidence interval when the sample is restricted to properties selling for at least \$ 250,000 (column (2)) only.

Results for the impact of downward risk-reclassification are presented in columns (3) and (4). I find that properties remaining in the floodplain experience a price increase of 15.6% ($exp(0.145)$) between the first and second sale. This coefficient is not statistically significant for *InOut* properties when estimated on the full sample, but it is 10.8 percentage points ($exp(0.103^{**})$ when estimated on properties above \$250,000.

Thus, I find that (1) upward risk-reclassification has a negative effect on real estate prices, but this effect is precisely estimated for properties above \$250,000 and that (2) downward risk-reclassification has a positive impact on real estate prices. These estimates are consistent with the real estate market capitalizing at least part of insurance premiums. Properties experiencing downward risk-reclassification appreciate because homeowners do not have to account for future flood insurance premiums (or much lower). That I am only able to estimate precisely the impact of upward and downward risk-reclassification on properties above \$250,000 should not be surprising. As mentioned above, the National Flood Hazard Insurance Program maximum deductible is \$250,000 for single-family and two- to four -dwelling residences. In addition to insurance premiums, homeowners of houses valued above \$ 250,000 concerned about flood risk also discount their potential net loss in case of a flood.

4.2.3 The role of awareness

A general approach on awareness and risk reclassification

If flood risk discount on property prices was largely driven by awareness, then we should observe larger discounts on the prices of properties in the floodplain when flood risk awareness is high. I present the results of the difference-in-differences model with fixed effects augmented with measures of awareness estimated on the panel of repeated sales in Table 12. Specifications are similar to columns (1) and (2) of Table 11.

Table 12: The Role of Awareness in Risk Reclassification Impact on House Prices
Main

	Dependent variable:			
	Ln Price			
	(1)	(2)	(3)	(4)
Post Update	0.182***	0.016**	0.181***	0.014**
	(0.007)	(0.007)	(0.007)	(0.007)
High Awareness Month			0.003	−0.012**
			(0.005)	(0.005)
Post Update * OutIn	0.034	−0.007	−0.014	−0.036
	(0.023)	(0.026)	(0.022)	(0.022)
Post Update * OutIn * High Awareness DMA	−0.141***	−0.072**		
	(0.039)	(0.035)		
Post Update * OutIn * High Awareness Month			−0.008	−0.019
			(0.042)	(0.036)
Sample	OutOut, OutIn	OutOut, OutIn	OutOut, OutIn	OutOut, OutIn
Minimum Transaction Price		250,000		250,000
Ref cat	OutOut	OutOut	OutOut	OutOut
Year dummy	Y	Y	Y	Y
Month dummy	Y	Y	Y	Y
Property FE	Y	Y	Y	Y
R2	0.872	0.89	0.871	0.89
Within R2	0.056	0.061	0.055	0.061
Observations	45,002	10,918	45,002	10,918

*p<0.1; **p<0.05; ***p<0.01

This table presents the results of the unit fixed effect model presented in equation (3) augmented with measures of awareness on the panel dataset of repeated sales restricted to *OutOut* and *OutIn* properties. Real estate transaction data and house characteristics are from Zillow. Floodplain data are from FEMA. I trim the real estate dataset to exclude the bottom and top 1% sales amount. In columns (1) and (2), *High Awareness DMA* is a dummy equal to 1 if average Google Index in the Designated Market Area (DMA) where is located the property is above or equal to the median of Google Index averages. In column (3) and (4), *High Awareness Month* is a dummy equal to 1 if the monthly-DMA Google Index is above the 75th percentile of the sample's monthly-DMA indexes. All specifications include property, fixed effects as well as year and month dummies. Standard errors are not clustered and presented below estimates.

In columns (1) and (2) I distinguish between low and high awareness DMAs (as defined according to their position compared to the median at the DMA level, averaged across all periods). Properties in the *OutOut* category sold on average 20% ($exp(0.182)$) higher on the second transaction (post update) than on the first one (pre update). *OutIn* properties in low awareness DMAs did not experience significantly different price trends. The price increase of *OutIn* properties in high awareness DMAs was 15.8 percentage points lower than for *OutOut* properties ($exp(0.182) - exp(0.182 - 0.141)$).

While this result does not inform us about the underlying sources of heterogeneity between regions, it is consistent with Bernstein et al. (2017), who find that sea level rise affects prices of non-owner occupied properties only in counties with a high enough *Worried* score on the Yale Climate Opinion Maps.

In columns (3) and (4) I study the effect of high awareness months on the transaction price of *OutIn* properties after the risk-reclassification. A high awareness month is defined as a month with Google Index above the 75 percentile of the 1,584 monthly-DMA indexes[27]. *OutIn* properties valued above \$250,000 sold on average at a 2.2% discount ($exp(0.014 - 0.036)$) after getting in the floodplain. This coefficient is not statistically different for high awareness months[28].

Eliciting a candidate mechanism: focus on the flood map updating process

The measure of flood risk awareness built in this paper is based on Google searches including flood risk related terms[29] and might therefore capture changes in risk awareness not directly relevant for the local real estate market. Figure 10 shows for example a peak in Southern Carolina Google Searches for *Flood* around September 2008, the month of Hurricane Ike, a \$30 billion damage disaster but which did not touch any of the states covered in this analysis. To address this issue I focus on the map updating process itself to show how a better targeting of critical moments gives insight on people's risk pricing process. The map updating timeline is summarized in figure 6.

I study how does the price of *OutIn* properties evolve along this timeline. Because of the small number of properties in my repeated sales panel sold in each of these periods, I move back to the hedonic specification presented in Equation (2)[30] which I run on all the *OutIn* properties of my sample (15,299 transactions) in columns (1)-(3) and on *OutIn* properties sold for more than \$250,000 in columns (4)-(6). As explained above, \$250,000 is the maximum flood insurance deductible for one-to-four family dwellings. I include two distinct measures of awareness. *High Awareness Month* is a dummy equal to 1 if the (main)Google Index is above the 75th percentile of the sample's monthly-DMA indexes. In addition, I define *High Awareness Month (b)* according to the same criteria, but using the *Google Index* built using the word *Flood* only.

[27] 12 DMAs across 11 years.

[28] $-0.036(0.022)$ is not statistically significant at traditional confidence intervals in 12 but it is significant when the model is ran without the *Post Update * OutIn * High Awareness Month* interaction (-0.043^{**} (0.018)).

[29] See Section 2.4 and Appendix A.5 for more details.

[30] Unlike results presented in Table 8, Census Tract T and Year Y fixed effects are not interacted together, as this would capture the monthly DMA variations of the variable *High Awareness Month*.

Table 13: Event Study: The role of awareness during the updating process

	Dependent variable:					
	Ln Price					
	(1)	(2)	(3)	(4)	(5)	(6)
Pre Update Process * High Awareness Month		−0.001			−0.004	
		(0.022)			(0.036)	
Pre Update Process * High Awareness Month (b)			−0.001			−0.015
			(0.024)			(0.027)
Preliminary BFE	−0.039	−0.042	−0.022	−0.065	−0.051	−0.074
	(0.094)	(0.106)	(0.094)	(0.084)	(0.085)	(0.086)
Preliminary BFE * High Awareness Month		−0.013			−0.113	
		(0.133)			(0.157)	
Preliminary BFE * High Awareness Month (b)			−0.293			−0.412***
			(0.399)			(0.089)
Appeal Period	−0.049	−0.040	−0.013	−0.001	0.017	0.024
	(0.117)	(0.160)	(0.157)	(0.082)	(0.090)	(0.108)
Appeal Period * High Awareness Month		−0.041			−0.095	
		(0.233)			(0.126)	
Appeal Period * High Awareness Month (b)			−0.117			−0.122
			(0.196)			(0.162)
Map Pending	−0.041	−0.077	−0.049	−0.042	−0.045	−0.092
	(0.084)	(0.096)	(0.100)	(0.091)	(0.085)	(0.090)
Map Pending * High Awareness Month		0.152			0.018	
		(0.152)			(0.114)	
Map Pending * High Awareness Month (b)			0.028			0.099
			(0.113)			(0.128)
Map Update Month	−0.016	−0.022	−0.020	−0.012	−0.017	−0.023
	(0.055)	(0.053)	(0.054)	(0.055)	(0.062)	(0.061)
Map Update Month * High Awareness Month		0.020			0.007	
		(0.030)			(0.028)	
Map Update Month * High Awareness Month (b)			0.022			0.018
			(0.026)			(0.031)
Post Update	0.156	0.154	0.151	0.019	0.019	0.056
	(0.104)	(0.110)	(0.193)	(0.248)	(0.248)	(0.121)
Post Update * High Awareness Month		0.065				
		(0.211)			(0.000)	
Post Update * High Awareness Month (b)			−0.001			−0.062
			(0.244)			(0.401)
Sample	OutIn	OutIn	OutIn	OutIn	OutIn	OutIn
Minimum Transaction Price				250,000	250,000	250,000
Size Controls	Y	Y	Y	Y	Y	Y
T*B*C FE	Y	Y	Y	Y	Y	Y
Y*B*C FE	Y	Y	Y	Y	Y	Y
R2	0.79	0.79	0.79	0.755	0.755	0.756
Within R2	0.026	0.027	0.027	0.061	0.061	0.062
Observations	15,299	15,299	15,299	5,984	5,984	5,984

*p<0.1; **p<0.05; ***p<0.01

This table presents the results of the event study on the role of awareness during the flood map updating process (see figure 6). Real estate transaction data and house characteristics are from Zillow. Floodplain data are from FEMA. I trim the real estate dataset to exclude the bottom and top 1% sales amount. The sample is then restricted to properties experiencing risk reclassification after their community's map update (*OutIn* properties). *High Awareness Month* is a dummy equal to 1 if the main monthly-DMA Google Index is above the 75th percentile of the sample's monthly-DMA indexes. *High Awareness Month (b)* is a dummy equal to 1 if the monthly-DMA Google Index based on the term "*Flood*" only is above the 75th percentile of the sample's monthly indexes for this index.

The only statistically significant coefficient is for the interaction between *Preliminary BFE* and *High Awareness Month* for properties valued more than \$250,000 when awareness is defined as using the alternative awareness index (only built of Google searches containing the word *Flood*). This coefficient is not statistically significant when awareness is defined by my main awareness measure (combining the aforementioned index and a broader one, as defined in Section 2.4). A possible explanation for this discrepancy is that this hybrid index captures searches reflecting people awareness of risk, but that they do not translate it into beliefs if they do not associate it with "*Flood*".

This result can be interpreted in light of what happens in each of these periods. After the preliminary base flood elevations (BFE) are released by FEMA, the agency must publish the notice in at least 2 local newspapers. Around this time, public meetings and open houses are also hold, as emphasized by Shawnee County Kansas documentation about their local 2010 flood map update:

> "*Sometime between January and March 2010, is when it is anticipated FEMA will publish the preliminary Base Flood Elevations (BFEs) in the Federal Register [...] Community officials have determined to hold no less than six (6) public meetings/open houses beginning January 2010. Questions generally range in topic from flood insurance and building requirements to mitigation opportunities and map changes.*"[31]

This finding echoes estimates presented in Table 11, when I showed that the estimate for *Post Update * OutIn* was negative and statistically significant (-0.045^{**}) when estimated on properties valued above \$250,000.

I discuss a candidate mechanism in Table 14. Column (1) confirms that in the *Map Pending* period, the *Preliminary BFE* period and on the month of the effective map update, more articles related to flood insurance are published by local newspapers (coefficients of 0.264^{***}, 0.243^{**} and 0.376^{*} respectively). These coefficients are not statistically significant in specifications without county fixed effects (column (2)).

In columns (3)-(4), Google Index is regressed on the different map update periods interacted with the monthly number of insurance related articles (weighted by their sentiment score) published in local newspapers. This makes the link between the increase in insurance articles published in these periods and the negative coefficient of -0.412 for *Preliminary BFE * High Awareness Month* estimated in column (6) of Table 13.

In columns (5)-(6), the alternative measure of awareness, *Google Index (b)* is regressed on the different map update periods interacted with the monthly number of insurance related articles (weighted by their sentiment score) published in local newspapers.

[31] Source : `http://snmapmod.snco.us/fmm/document/09-flood-map-production-process.pdf`.

Table 14: Event Study - Impact of the map updating process on local newspaper articles and Awareness

	Dependent variable:					
	Insurance Articles (scaled)		Google Index		Google Index (2)	
	(1)	(2)	(3)	(4)	(5)	(6)
Preliminary BFE	0.243** (0.108)	0.047 (0.172)	−0.131 (0.094)	−0.385*** (0.114)	−0.217* (0.117)	−0.421*** (0.129)
Prelim * Insurance Art.			−0.014 (0.016)	−0.002 (0.023)	−0.019 (0.026)	−0.010 (0.027)
Prelim * Insurance Art. * Score			−0.004** (0.002)	−0.004** (0.002)	−0.002 (0.002)	−0.002 (0.002)
Appeal	0.058 (0.108)	−0.134 (0.149)	−0.207* (0.110)	−0.454*** (0.119)	−0.453*** (0.133)	−0.662*** (0.130)
Appeal * Insurance Art.			−0.015 (0.023)	−0.006 (0.022)	−0.019 (0.025)	−0.012 (0.023)
Appeal * Insurance Art. * Score			−0.003 (0.003)	−0.001 (0.002)	0.001 (0.003)	0.002 (0.002)
Map Pending	0.264*** (0.101)	0.064 (0.145)	−0.415*** (0.104)	−0.694*** (0.122)	−0.478*** (0.121)	−0.703*** (0.117)
Pending * Insurance Art.			0.019 (0.018)	0.039* (0.020)	−0.004 (0.023)	0.008 (0.020)
Pending * Insurance Art. * Score			0.0002 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
Update	0.376* (0.200)	0.126 (0.236)	−0.170 (0.185)	−0.435** (0.186)	−0.188 (0.182)	−0.397** (0.191)
Update * Insurance Art.			0.032 (0.093)	0.043 (0.100)	−0.007 (0.100)	−0.004 (0.099)
Update * Insurance Art. * Score			−0.027*** (0.004)	−0.024*** (0.005)	−0.028*** (0.003)	−0.026*** (0.004)
Flood controls	Y	Y	Y	Y	Y	Y
Location FE	County		County		County	
Year FE	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y
R2	0.286	0.034	0.305	0.277	0.293	0.273
Within R2	0.033	0.025	0.19	0.192	0.186	0.189
Observations	22,704	22,704	22,704	22,704	22,704	22,704

*p<0.1; **p<0.05; ***p<0.01

This table presents the results of the event study estimating the impact of each of the map updating period on the number of insurance articles in columns (1) and (2) and on the main *Google Index* in columns (3) and (4), and on *Google Index (b)* in columns (5) and (6). The models are run on the local newspaper panel of the 172 newspapers with available circulation estimates between 2006 and 2016. Observations are at the newspaper-month level.
The number of insurance articles is scaled from 0 to 100 in a similar way to the *Google Index*. *Preliminary BFE* is a dummy equal to 1 if the property was sold between 12 and 10 months before the updated map became effective ; *Appeal Period* is a dummy equal to 1 if the property was sold between 9 and 7 months before the map became effective on the *Map Update Month* and *Map Pending* is a dummy equal to 1 if the property was sold between 6 and 1 months before the *Map Update Month*.
Standard errors are clustered at the Newspaper level and presented next to estimates. In columns (3) and (4), observations are weighted by newspaper circulation.
The lower the insurance articles *Score*, the more negative the articles. The variable *Score* go from -108 to 50.

Controlling for County fixed effects, I estimate coefficients of -0.004^{**} and -0.027^{***} for the interactions between the number of insurance articles weighted by their sentiment score, the *Preliminary BFE* the *Map Update Month* periods respectively. For example, for a given number of insurance articles published by a newspaper at each stage of the map update process, a 10 points decrease in their average score will lead to an increase in *Google Index* of 0.04 and 0.27 points. The number of insurance articles itself is only relevant in the map pending period when not controling for county fixed effects (0.039^{*}).

These results suggest that insurance concerns are key in homeowners internalization of flood risk in their buying decisions. More flood insurance-related articles are released all along the map update process, and I identify that these publications, when associated with a negative sentiment score, translate into increased awareness in the *Preliminary BFE* and *Month Update* periods.

However, future work will make sure to investigate why no significant effect of *Prelim * Insurance Art * Score* can be estimated when the alternative measure of awareness, *Google Index (b)* is used, while it is with this index only that the impact on real estate prices (Table 13) can be estimated precisely.

4.3 Robustness Checks

I run a series of robustness checks to address possible remaining concerns for the three models presented above.

First, I account for the fact that houses in the floodplain are more likely to have been flooded, and flooded at a greater intensity; than properties outside the floodplain. To do so, I exclude properties located in counties mentioned in a Flood or Hurricane Presidential Disaster Declaration [32] before the property's sale. In Appendix B.2, Tables 20, 21 and 22 present the corresponding for estimates for the hedonic analysis, the simple unit fixed effect model and the awareness-augmented unit fixed effect modelrespectively.

The *Floodplain* coefficients of the hedonic analysis estimated on non-flooded counties (Table B.2) are all non statistically significant, even for distance-to-coast cutoffs greater than 0.5 miles. The absence of statistical significance compared to the main results presented in Table 9 may mean that in non-flooded counties, the premium associated with local amenities potentially captured by the *Floodplain* coefficient is less strong and simply balances the discount associated with flood risk. This is likely to be true if non-flooded counties are not coastal counties.

The estimates for risk-reclassification in Table 21 are consistent with the main estimates (Table 11) and of a slightly smaller magnitude (-0.037^{*} versus -0.045^{**} for upward risk-reclassification of properties above \$250,000).

[32] As of today, my flood history database only contains floods between, 2006 and 2016 and are therefore only excluded properties that were flooded before, starting in 2006. Future work will need to account for floods before the 2006-2016 period.

In Table 22, the coefficients for *Post Update * OutIn * High Awareness DMA* drops and only remains significant at the 90% confidence interval when estimated on the full sample (0.109^{*} versus -0.141^{***}). It is not significant anymore for the sample of > \$250,000 properties ($-0.071(0.049)$ versus 0.072^{**}).

These results suggest that local flood history impacts real estate participants' expectation about future floods and, more precisely, that in counties that have not experienced natural disasters in recent years, flood risk as measured by the SFHA is less credible to homeowners.
Second, I investigate how my flood risk reclassification estimates change when restricting my sample to coastal properties.

Third, I have chosen before to run the risk-reclassification difference-in-differences model on the panel of repeated sales restricted to properties valued above \$250,000. Figure 13 graphs the evolution of the *Post Update * OutIn* and *Post Update * InOut* when moving up this threshold between \$50,000 and \$400,000, which corresponds to the 90^{th} percentile of the transaction price of properties in my repeated sales panel. The *Post Update * InOut* coefficient is always positive and overall significant, suggesting that getting out of the floodplain is a positive signal on the property's value for all properties. The *Post Update * OutIn* coefficient is close to 0 until the \$180,000 threshold and is negative and statistically significant at the 90% confidence interval starting around \$220,000. The small size of the sample of properties above \$320,000 might explain why the coefficient's significant level falls while remaining negative when moving the price lower bound up.

The absence of a clear coefficient change at \$250,000 goes against the use of a random discontinuity design model (RDD) and is consistent with the fact that transaction price might be different (lower) to buyers' valuation of the property, leading them to potentially insure it above its transaction value.

Fourth, I check that my results about the impact of awareness on the degree of flood risk capitalization on the housing market are robust to alternative definitions of *High Awareness*. Table 23 compares the results of the map update event study when high awareness is defined as being above or equal the mean (rather than the third quartile) of the full sample's *Google Index* (Google Index equal to 4.6). The coefficient for *Preliminary BFE * High Awareness Month (2b)* is negative and statistically significant for the second Google index definition, and negative but not statistically significant for the first one.

5 Discussion

5.1 Concluding Remarks

There is no consensus in the empirical literature on the impact of objective flood risk on real estate prices. In this paper I have tried to answer two questions : *When are people aware of flood risk? When does the real estate market capitalize this risk?* Taken together, they offer new insights for this absence of agreement.

First, I have emphasized that flood risk was capitalized heterogeneously on the housing market and that this heterogeneity had three potential complementary explanation. I have been able to consistently estimate flood risk associated discounts for houses above $250,000 only. A first explanation for this finding can be found in the behavioral economic literature: Handel and Schwartzstein (2018) review situations where people fail to use available information to make decisions and insist on the existence of "mental gaps". Limited financial, scientific literacy or a high discount rate, which are factors highly correlated with people's socioeconomic status, could lead less educated people to under estimate flood risks. A second explanation is to find on the flood insurance side, as $250,000 is also the maximum deductible under the National Flood Insurance Program. Observing flood risk discounts for properties above this threshold would be consistent with homeowners pricing damage not covered by insurance. Insurance concerns also seem to drive flood risk awareness up. Besides, I have found that articles about flood insurance featuring a low sentiment score (negative) amplified the impact of flood map updates on awareness.

Does this imply that awareness is a friction-less channel between information shocks and flood risk beliefs? No, it does not: real estate prices seem to capitalize flood risk more when transactions happen around a small subset of all related information shocks only. Flood experience has a positive impact on the degree to which awareness translates into beliefs, suggesting that, in this case, not only people know this risk exists, but they believe it could happen to them.

5.2 Possible Extensions

This paper has focused on estimating the impact of flood risk and on understanding whether updates in flood risk beliefs systematically followed improvements in flood risk awareness. A number of points would deserve to be investigated in the future.

Additional Robustness Checks

For the sake of time, only a limited number of robustness checks are presented here, but future work should investigate at least three other potential concerns.

First, while all covariate coefficients of the hedonic analysis have the expected sign, the magnitude for *Ln (Square Feet)* is significantly larger than in comparable studies, especially for the 0.25 miles cutoff. I will make sure to investigate this issue and delve into two potential sources for these differences, namely Zillow database's highly state-specific property size sources and consequences of my specifications with interacted fixed effects[33].

Second, I believe I could improve the hedonic model by adding controls to better capture topological and location specific amenity values such as a property's elevation (Bernstein et al. (2017)), proximity to parks, Starbucks (Troy and Romm (2004)), by moving to a more precise distance-to-coast measure, and by adding blockgroup or block specific neighborhood characteristics[34]

Extending the event study to post flood map update periods

Besides, I would be interested in extending the event study presented in Section 4.2.3 to post flood map update periods. More specifically, I would like to study whether the estimated effect of flood-risk reclassification on property prices is long lasting or not. In a study on flood insurance take up after a neighboring flood, Gallagher (2014) estimates that homeowners' reaction over time is consistent with a Bayesian learning model with forgetting.

Improving the scope of information shocks

Extending the range of information shocks by adding other information sources could be done in two main ways. First, following the methodology developed by Snyder and Stromberg (2010) and exploited by Gallagher (2014) to study flood insurance take up, for analyzing the impact of local TV media and their content.

Furthermore, future work should think more about how to measure the intensity of public discussions, meetings, and official declarations about flood risk. Investigating this dimension of heterogeneity could bring additional light on how a community's dynamism, structure and political life matter for the internalization of environmental and economic issues like flood risk by the population.

Implications for Structural Models

Last but not least, these findings have potential implications for structural housing models. Gibson et al. (2017) developed a theoretical structural model to explain changes in house prices after the Hurricane Sandy, an insurance reform and flood map updates. Their goal was to isolate the role of belief update. One could imagine augmenting this model with awareness and media pressure heterogeneity in order to explain better heterogeneity of flood risk pricing observed in the data. More generally, this paper's findings help to understand how households update their beliefs. Modeling this updating process is a key dimension of structural models, and at the frontier of research.

[33] Bernstein et al. (2017) find lower magnitude coefficients but (1) the log of their sample's average building size is lower (3.23 vs 7.26) and (2) they are able to control for 4th order polynomial distance to the beach, in addition to distance-to-coast buffers fixed effects thanks to very granular distance measures.

[34] The main specification has a Census Tract * Month fixed effect. Adding annual demographic controls at the county level like median income, median age or share of hispanic population would be redundant.

References

A. Bernstein, M. Gustafson, and R. Lewis. Disaster on the Horizon: The Price Effect of Sea Level Rise. *SSRN Electronic Journal*, 2017.

O. Bin and C. E. Landry. Changes in implicit flood risk premiums: Empirical evidence from the housing market. *Journal of Environmental Economics and Management*, 65(3):361–376, May 2013.

O. Bin, J. B. Kruse, and C. E. Landry. Flood Hazards, Insurance Rates, and Amenities: Evidence From the Coastal Housing Market. *Journal of Risk & Insurance*, 75(1):63–82, Mar. 2008.

H. Choi and H. Varian. Predicting the Present with Google Trends. *Economic Record*, 88(s1):2–9, 2012.

M. Crowell, K. Coulton, C. Johnson, J. Westcott, D. Bellomo, S. Edelman, and E. Hirsch. An estimate of the us population living in 100-year coastal flood hazard areas. *Journal of Coastal Research - J COASTAL RES*, 26, 03 2010.

FEMA. Policy and claim statistics for flood insurance. 2017. URL `https:www.fema.gov/policy-claim-statistics-flood-insurance`.

J. Freybote and E. Fruits. Perceived Environmental Risk, Media and Residential Sales Prices. *Journal of Real Estate Research*, 37(2):217–244, 2015.

J. Gallagher. Learning about an Infrequent Event: Evidence from Flood Insurance Take-Up in the United States. *American Economic Journal: Applied Economics*, 6(3):206–233, July 2014. ISSN 1945-7782.

M. Gentzkow, J. M. Shapiro, and M. Sinkinson. The effect of newspaper entry and exit on electoral politics. *American Economic Review*, 101(7):2980–3018, December 2011. doi: 10.1257/aer.101.7.2980.

A. S. Gerber, D. Karlan, and D. Bergan. Does the media matter? a field experiment measuring the effect of newspapers on voting behavior and political opinions. *American Economic Journal: Applied Economics*, 1(2):35–52, April 2009. doi: 10.1257/app.1.2.35.

M. Gibson, J. T. Mullins, and A. Hill. Climate change and flood risk: Evidence from New York real estate. page 66, 2017.

D. G. Hallstrom and V. K. Smith. Market responses to hurricanes. *Journal of Environmental Economics and Management*, 50(3):541–561, Nov. 2005.

B. Handel and J. Schwartzstein. Frictions or Mental Gaps: What's Behind the Information We (Don't) Use and When Do We Care? *Journal of Economic Perspectives*, 32(1):155–178, Feb. 2018.

D. M. Harrison, G. T. Smersh, and J. Arthur L. Schwartz. Environmental Determinants of Housing Prices: The Impact of Flood Zone Status. *Journal of Real Estate Research*, 21(1/2):3–20, 2001.

A. Indaco, F. Ortega, and S. Taspnar. The Eects of Flood Insurance on Housing Markets. *IZA DP No. 11810*, page 40, 2018.

M. Kahn and M. Kotchen. Environmental Concern and the Business Cycle: The Chilling Effect of Recession. Technical Report w16241, National Bureau of Economic Research, Cambridge, MA, July 2010.

C. Kousky. Learning from Extreme Events: Risk Perceptions after the Flood. *Land Economics*, 86 (3):395–422, 2010.

H. Kunreuther. Reflections on U.S. Disaster Insurance Policy for the 21st Century. Technical Report w12449, National Bureau of Economic Research, Cambridge, MA, Aug. 2006.

H. Kunreuther, R. Meyer, and E. Michel-Kerjan. Overcoming decision biases to reduce losses from natural catastrophes. *The Behavioral Foundations of Public Policy*, 09 2009.

C. Lang and J. D. Ryder. The effect of tropical cyclones on climate change engagement. *J.D. Climatic Change*, 2006.

J. McCluskey and G. Rausser. Estimation of perceived risk and its effect on property values. *Land Economics*, 77(1):42–55, 2001.

D. Moore. Nfip vs. private flood insurance. 2017. URL `http://www.davemooreinsurance.com/news-and-resources/nfip-vs-private-flood-insurance/`.

S. Pralle. Drawing Lines: FEMA and the politics of mapping flood zones, 2017.

S. Rosen. Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1):34–55, 1974.

J. M. Snyder and D. Stromberg. Press coverage and political accountability. *Journal of Political Economy*, 118(2):355–408, 2010.

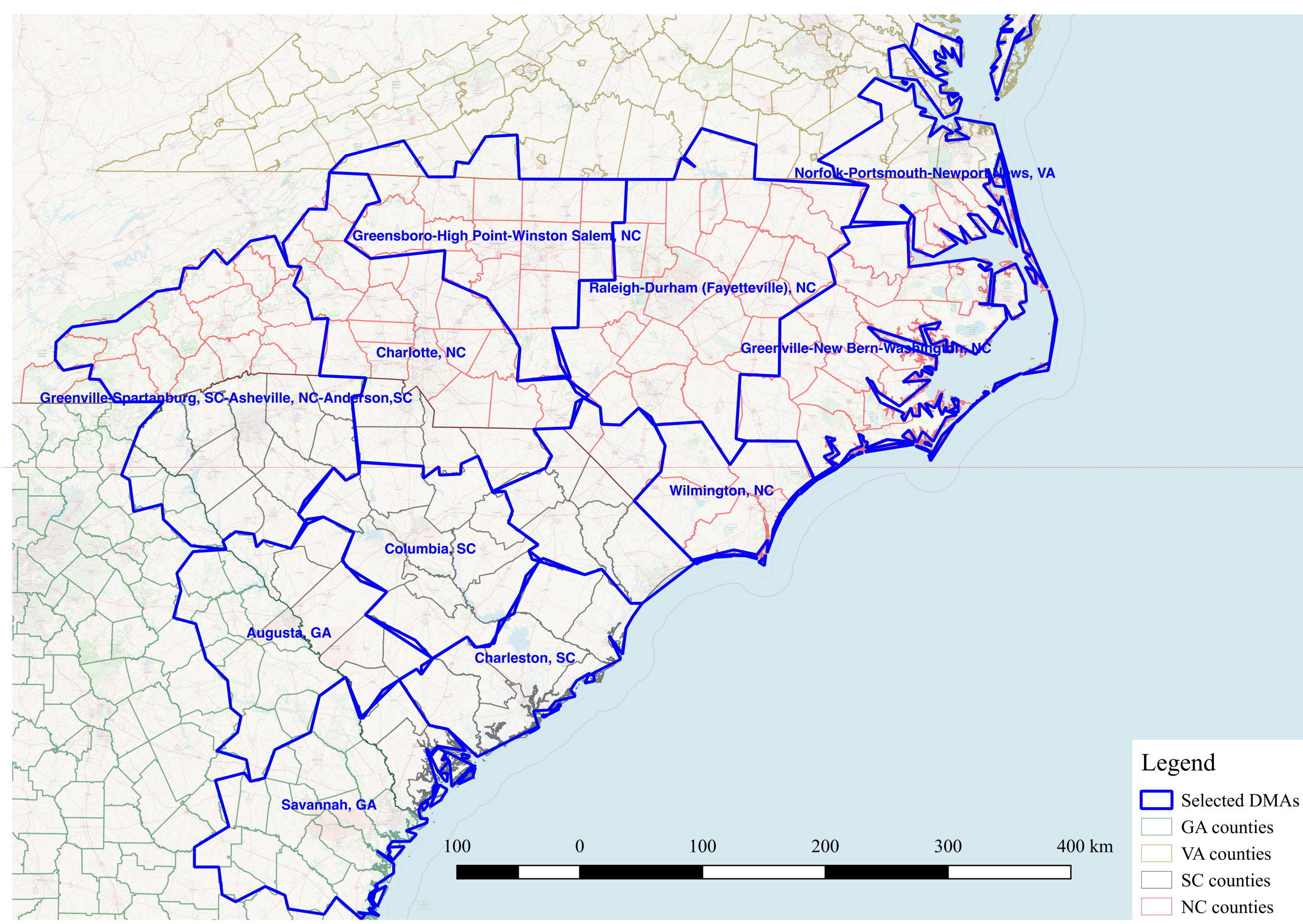
A. Troy and J. Romm. Assessing the price effects of flood hazard disclosure under the California natural hazard disclosure law (AB 1195). *Journal of Environmental Planning and Management*, 47(1):137–162, Jan. 2004.

A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185 (4157):1124–1131, 1974.

A Data

A.1 Geographical Coverage

Figure 7: Mapping of counties into DMAs



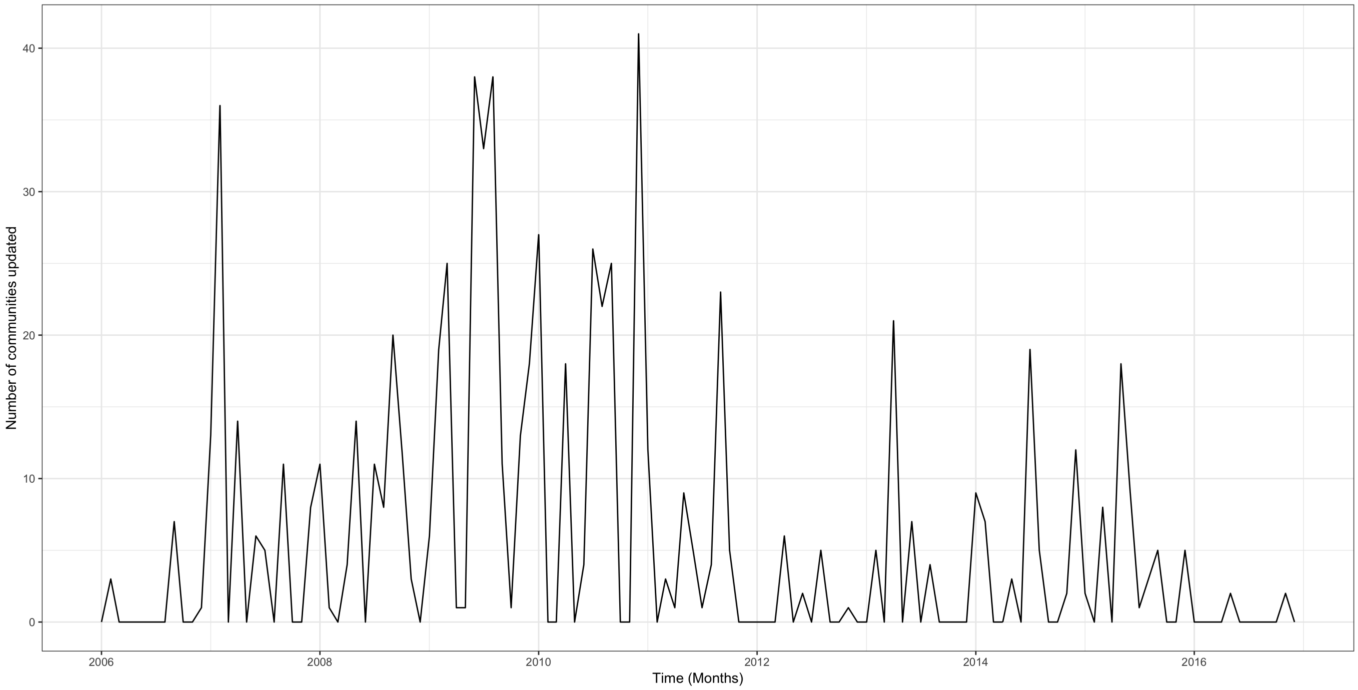
This map shows the counties covered by each of the 12 DMAs studied in this paper. Each DMA contains between 5 and 29 counties. The corresponding states are North Carolina (NC), South Carolina (SC), Virginia (VA) and Georgia (GA).

A.2 Housing Data

Are coded as residential properties, properties within any of the following categories on the Transaction table: *Apartment Building (AP)*, *Condominium (CD)*, *Cooperative (CP)*, *Mobile Home (MB)*, *Multi-Family Dwelling (MF)*, *Manufactured Home (MH)*, *Mixed Use (MX)*, *New Construction (NW)*, *Planned Unit Development (PD)*, *Residential (RR)*, *Single Family Residence (SR)*, *Unimproved Land/Lot (UL)*, *Vacant Land/Lot (VL)*. If missing, filtering is based on the residential codes on the Assessment table.

A.3 Flood risk data

Figure 8: Flood Map Updates - Community level



This graph displays the number of Communities (identified by a Community Identification Number (CID)), which have seen their Flood Insurance Rate Map (FIRM) updated in a given month. A community corresponds to a small town or a city's neighborhood. PDD data is extracted from FEMA's website.

A.4 Flood History

A.5 Google Search Index

Comparison with *Neutral* search terms

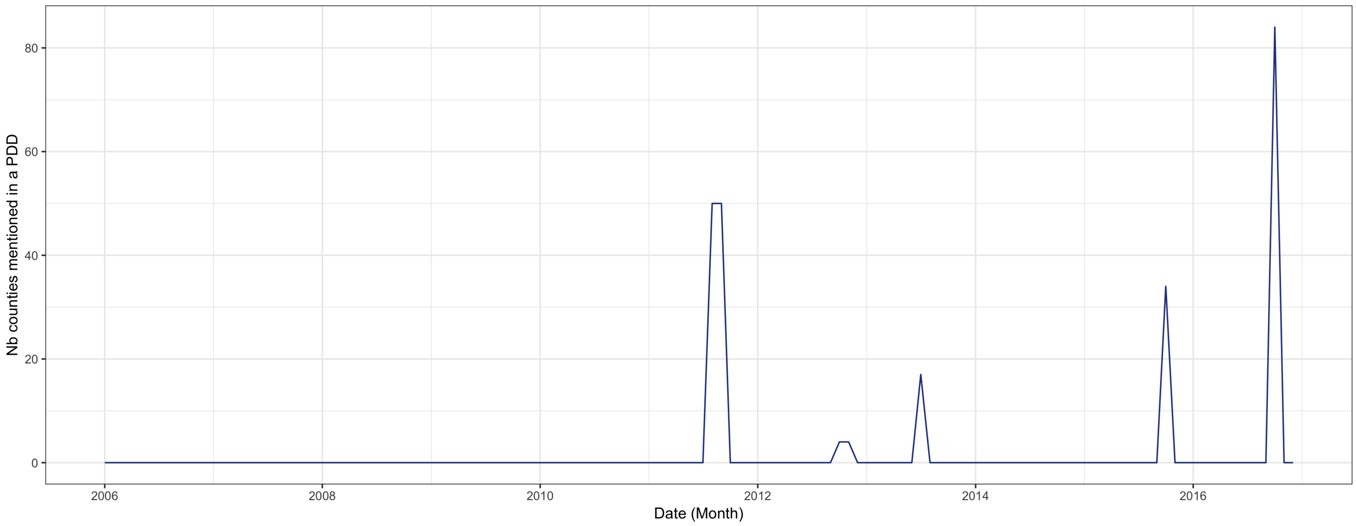
A.6 Local News

Sample building

Article Categories

Monthly articles count

Figure 9: Total number of Presidential Disaster Declarations

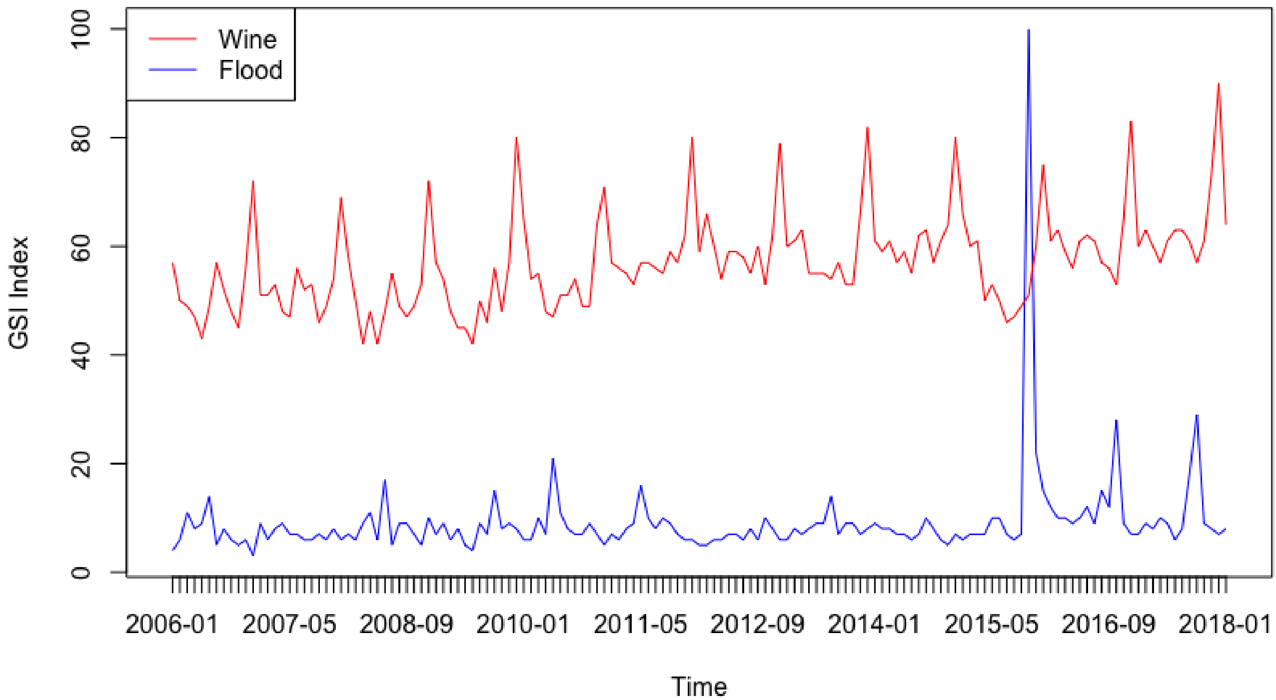


This graph shows the number of counties that filled a Presidential Disaster Declaration (PDD) for a *Flood* or a *Hurricane* in the 12 DMAs covered by the analysis between January 2006 and December 2016. PDD data is extracted from FEMA's website

B Robustness Checks

B.1 What Drives Awareness ?

Figure 10: Google Search Index Comparison - South Carolina



This graph compares the evolution of two Google searches between 2006 and 2018 in South Carolina according to Google's relative search index. Data is from Google Trends. Unlike the search for *Wine*, whose popularity follows seasonal trends (note the December peaks), the search for *Flood* is not subject to the same seasonality. Instead, it is low and stable. Its peaks correspond to major flood events (see for example the North American Storm Complex in October 2015 and Hurricane Matthew in July 2016).

Table 15: News articles chosen keywords

flood	*flood* [and] *risk*
flood [and] *map*	*flood* [and] *mitigation*
flood [and] *surge*	*sea* [and] *level* [and] *rise*
floodplain	*Grimm Waters*
flood [and] *insurance*	*flood* [and] *hurricane*
National [and] *Flood* [and] *Hazard* [and] *Layer*	*Biggert Waters*
Flood [and] *Insurance* [and] *Rate* [and] *Map*	*flood* [and] *storm*
map [and] *modernization*	*flood* [and] *study*
flood [and] *hazard*	*FIRM* [and] *flood*
flood [and] *plain*	*floodway*
flood [and] *lines*	*risk of future flooding*
NFIP	

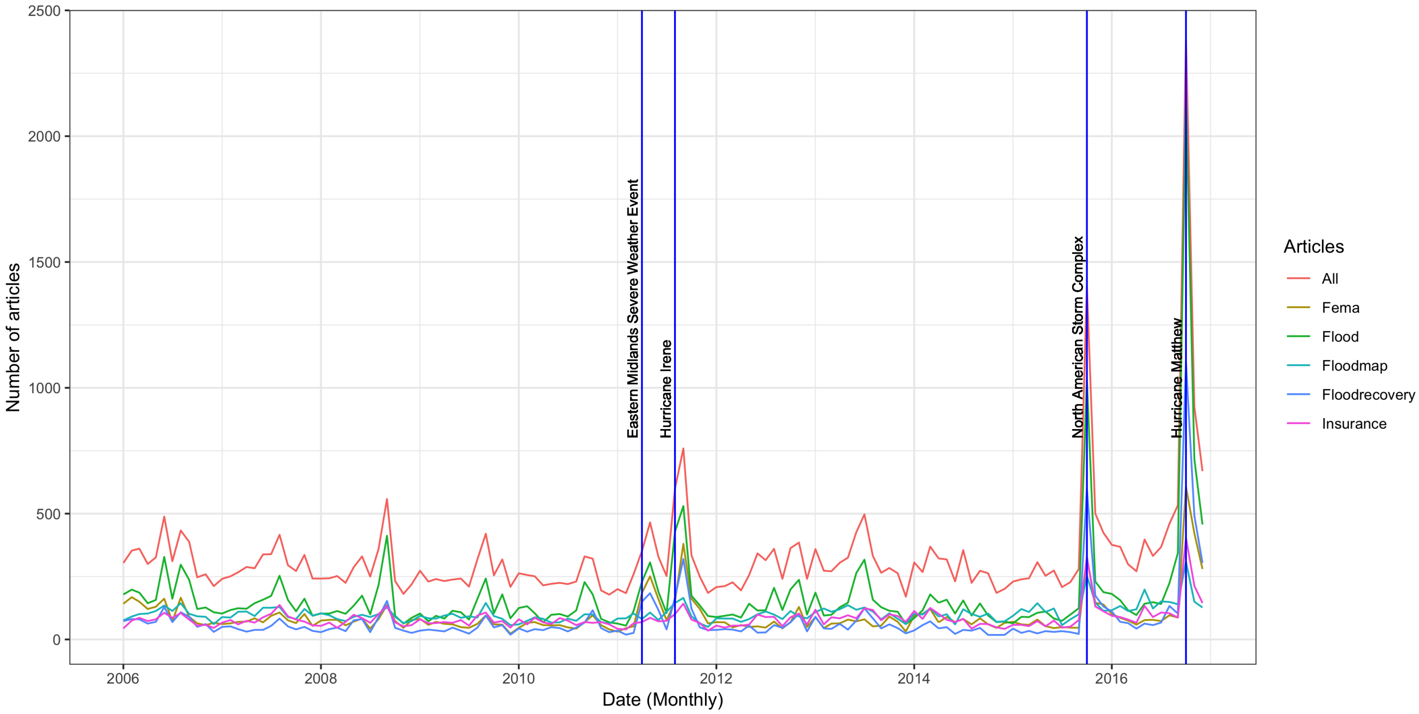
This table lists the terms used on the NewsBank plateform to constitute my local newspaper articles sample. All articles published by newspapers covering the 199 counties of interest were included in the sample if they contained at least of of these expressions.

Table 16: Word categories

Category	**Words**
INSURANCE	*Premium, Insurance, Rate, Ordinance, Flood insurance change, Insured, Coverage*
LAW	*Law, Congress, Regulation, Biggert Waters, Grimm Waters, Senate, Bill*
FEMA	*FEMA, Federal*
CHANGE	*Change, Update, New, Adapt, Reform, Modernization*
FLOOD MAP	*Flood map, Disclose flood settlement, Flood risk, Map, Flood planning, Flood control, Prevent, To cut flood damage, Mitigation plan*
MEETING	*Public event, Meeting, Conference, Open house, Panel*
FLOOD RECOVERY	*Recovery, Recovering, Recover, Damage, Repairs, Relief, Storm response, Rebuild, Victim, Debris removal*
FLOOD	*Emergency, Flooded, Rain, Katrina, Hurricane, Storm, Matthew, Tornado, Tsunami*
CLIMATE CHANGE	*Climate change, Global warming, Warmer, Wetter, Sea level rise, Carbon, Rising sea level, Scientist*
HOUSING	*Residents, Homeowners, House, Housing, Developers, Real estate*

This table lists the words and expressions falling into each of the 10 categories used to classify news articles.

Figure 11: Monthly Articles Count



This figures plots the monthly number of newspaper article releases between January 2006 and December 2016 for the 187 local newspapers in my panel. Articles are recovered from Newsbank. The total number of articles is plotted in red. Counts for five of the article categories constructed according to the rule described in Table 20 are also displayed on the graph. Peaks in article release correspond to major floods.

Table 17: The drivers of awareness
No County Fixed Effect

	Dependent variable:			
	Google Index			
	(1)	(2)	(3)	(4)
Flood	16.367***	14.758***		
	(1.225)	(1.394)		
Ln (Flood damage)		0.094		0.095
		(0.148)		(0.148)
Map Update	0.107	0.041	0.105	0.038
	(0.173)	(0.177)	(0.173)	(0.177)
Articles (weighted)		0.032		0.031
		(0.027)		(0.027)
Flood * Articles (weighted)		0.267**		
		(0.113)		
Ln (Flood damage) * Articles (weighted)			1.098***	0.997***
			(0.080)	(0.092)
Map Update * Articles (weighted)				0.016**
				(0.007)
County FE				
Year FE	Y	Y	Y	Y
Month FE	Y	Y	Y	Y
R2	0.249	0.258	0.257	0.264
Within R2	0.169	0.179	0.178	0.186
Observations	26,268	26,268	26,268	26,268

*p<0.1; **p<0.05; ***p<0.01

This table presents the ordinary least squares estimates corresponding to equation (1). The dataset covers 199 counties between 2006 and 2016 (132 months). Specifications are similar to Table 5 but no county fixed effect is included. Standard errors are presented below estimates and are clustered at the county level. Results are consistent with the estimates presented in Table 5.

Table 18: The drivers of awareness - Newspaper panel
No County Fixed Effect

	Dependent variable:			
	Google Index			
	(1)	(2)	(3)	(4)
Flood	12.307***		12.679***	
	(1.197)		(1.213)	
Map Update	0.368*	0.364*	0.385**	0.382**
	(0.191)	(0.191)	(0.168)	(0.168)
Articles	0.070***	0.069***		
	(0.016)	(0.016)		
Flood * Articles	0.299***			
	(0.081)			
Ln (Flood damage)		0.832***		0.856***
		(0.078)		(0.079)
Map Update * Articles	0.018	0.019		
	(0.134)	(0.134)		
Ln (Flood damage) * Articles		0.018***		
		(0.005)		
Flood Recovery Articles			0.097***	0.094***
			(0.025)	(0.025)
Insurance Articles			0.026	0.027*
			(0.016)	(0.016)
Insurance Articles * Score			−0.002	−0.002
			(0.002)	(0.002)
Flood * Flood Recovery Articles			0.235**	
			(0.105)	
Map Update * Insurance Articles			−0.003	−0.003
			(0.057)	(0.057)
Ln (Flood damage) * Flood Recovery Articles				0.014**
				(0.007)
Map Update * Insurance Articles * Score			−0.018***	−0.018***
			(0.006)	(0.006)
Location FE				
Year FE	Y	Y	Y	Y
Month FE	Y	Y	Y	Y
R2	0.297	0.303	0.305	0.34
Within R2	0.216	0.222	0.224	0.231
Observations	24,684	24,684	24,684	24,684

*p<0.1; **p<0.05; ***p<0.01

This table presents the results of equation (1) estimated on a monthly newsaper panel of 187 local newspaper between 2006 and 2016. Specifications are similar to Table 7 but no county fixed effect is included. Standard errors are presented below estimates and are clustered at the newspaper level. Results are consistent with the estimates presented in Table 7 .

Table 19: The drivers of awareness - Newspaper panel dataset
By Article category

	Dependent variable:			
	Google Index			
	(1)	(2)	(3)	(4)
Flood	12.335*** (1.213)	12.009*** (1.227)	12.232*** (1.205)	12.120*** (1.198)
Map Update	0.577*** (0.164)	0.594*** (0.165)	0.610*** (0.169)	0.559*** (0.160)
Flood Recovery Articles	0.116*** (0.026)	0.154*** (0.022)	0.218*** (0.042)	0.134*** (0.027)
Insurance Articles	0.025* (0.015)			
Insurance Articles * Insurance Score	−0.002 (0.002)			
Flood map Articles		−0.001 (0.009)		
Flood map Articles * Flood Map Score		−0.0003 (0.001)		
FEMA Articles			−0.063*** (0.020)	
FEMA Articles * FEMA Score			−0.001 (0.001)	
Law Articles				−0.002 (0.006)
Law Articles * Law Score				−0.002 (0.001)
Flood * Flood Recovery Articles	0.222** (0.104)	0.285*** (0.106)	0.237** (0.105)	0.268*** (0.095)
Map Update * Insurance Articles	0.007 (0.058)			
Map Update * Insurance Articles * Insurance Score	−0.021*** (0.005)			
Map Update * Flood map Articles		0.008 (0.052)		
Map Update * Flood map Articles * Flood map Score		−0.008 (0.006)		
Map Update * FEMA Articles			−0.033 (0.075)	
Map Update * FEMA Articles * FEMA Score			−0.006 (0.009)	
Map Update * Law Articles				0.018 (0.060)
Map Update * Law Articles * Law Score				−0.0002 (0.003)
Location FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Month FE	Y	Y	Y	Y
R2	0.335	0.332	0.333	0.333
Within R2	0.225	0.222	0.224	0.224
Observations	24,684	24,684	24,684	24,684

*p<0.1; **p<0.05; ***p<0.01

This table presents the results of equation (1) estimated on a monthly newsaper panel of 187 local newspaper between 2006 and 2016. Specifications are similar to Table 6 but no county fixed effect is included. Standard errors are presented below estimates and are clustered at the newspaper level. Results are consistent with the estimates presented in Table 6. In column (1) and (2), the coefficient for the interaction between *Ln (Flood damage* and *Articles* is significant at the 90% confidence interval.

B.2 Flood Risk and The Housing Market

Robustness to flood history

Table 20: The Impact of Floodplain Status on House Prices
Excluding Past Floods and Hurricanes

	Dependent variable:					
	Ln Price					
	(1)	(2)	(3)	(4)	(5)	(6)
Floodplain	0.125	0.084	0.035	0.043	0.047	0.037
	(0.126)	(0.076)	(0.066)	(0.058)	(0.042)	(0.034)
Size Controls	Y	Y	Y	Y	Y	Y
Age Controls						
T * L * D * B	Y	Y	Y	Y	Y	Y
T * L * D * B * BF * U						
Location FE	Tract	Tract	Tract	Tract	Tract	Tract
Maximum distance to the coast	0.25 mile	0.5 mile	1 mile	2 miles	5 miles	10 miles
R2	0.836	0.893	0.859	0.85	0.844	0.841
Within R2	0.062	0.029	0.081	0.091	0.087	0.087
Observations	1,180	3,481	11,217	24,624	54,522	68,628

*p<0.1; **p<0.05; ***p<0.01

This table presents the ordinary least squares estimate for the explanatory variable of interest, *Floodplain*, corresponding to equation (2) with samples restricted to properties within 0.25, 0.5, 1, 2, 5 and 10 miles from the coast. Real estate transaction data and house characteristics are from Zillow. Floodplain data are from FEMA. I trim the real estate dataset to exclude the bottom and top 1% sales amount. Unlike Table 9, transactions of properties located in counties mentioned in a Flood or Hurricane Presidential Disaster Declaration between 2006 and the month of the transaction are excluded.
All specifications include 4th order polynomials for building size as covariates as well as simple and interacted fixed effects for time, location, distance to the coast buffer (every 0.05 mile up to 0.5 mile, every 0.1 mile up to 10 miles) and number of bedrooms. Standard errors are presented below estimates and clustered at the location fixed effect level and at the source of the size data (*Total Building Area*, *Living Building Area*, *Heated Building Area*...).
Results are consistent to the ones presented in table 9 for the full sample.

Table 21: Risk Reclassification Impact on House Prices
Excluding Past Floods and Hurricanes

	Dependent variable:			
	Ln Price			
	(1)	(2)	(3)	(4)
Post Update	0.205***	0.023***	0.269***	0.013
	(0.007)	(0.008)	(0.058)	(0.089)
Post Update * OutIn	−0.001	−0.037*		
	(0.022)	(0.022)		
Post Update * InOut			0.013	0.113**
			(0.038)	(0.056)
Categories	OutOut, OutIn	OutOut, OutIn	InIn, InOut	InIn, InOut
Minimum Transaction Price		250,000		250,000
Ref cat	OutOut	OutOut	InIn	InIn
Year dummy	Y	Y	Y	Y
Month dummy	Y	Y	Y	Y
Property FE	Y	Y	Y	Y
R2	0.872	0.893	0.899	0.859
Within R2	0.065	0.062	0.127	0.262
Observations	39,842	9,582	802	230

*p<0.1; **p<0.05; ***p<0.01

This table presents the results of the unit fixed effect model summarized in equation (3) estimated on a panel of properties that were sold both before and after their community's map update. Real estate transaction data and house characteristics are from Zillow. Floodplain data are from FEMA. I trim the real estate dataset to exclude the bottom and top 1% sales amount. When the property was sold more than once before or after the update, only the closest sale to the update is considered. Properties sold in counties mentioned in a Flood or Hurricane Presidential Disaster Declaration before the second sale are excluded from the panel. In column (1) and (2), *OutOut* and *OutIn* are considered. The reference category is the *OutOut* properties sold before the update. In column (3) and (4), *InIn* properties are compared to *InOut* properties. All specifications include property fixed effects as well as year and month dummies. Unlike odd columns, results presented in even columns are estimated on a subset of the sample excluding properties sold for less than $250,000. Standard errors are presented below estimates and are not clustered.
While the magnitude of the *Post Update * OutIn* and *Post Update * InOut* coefficients is larger for columns (2) and (4), no interaction is statistically significant at conventional confidence intervals.

Table 22: The Role of Awareness in Risk Reclassification Impact on House Prices Excluding Past Floods and Hurricanes

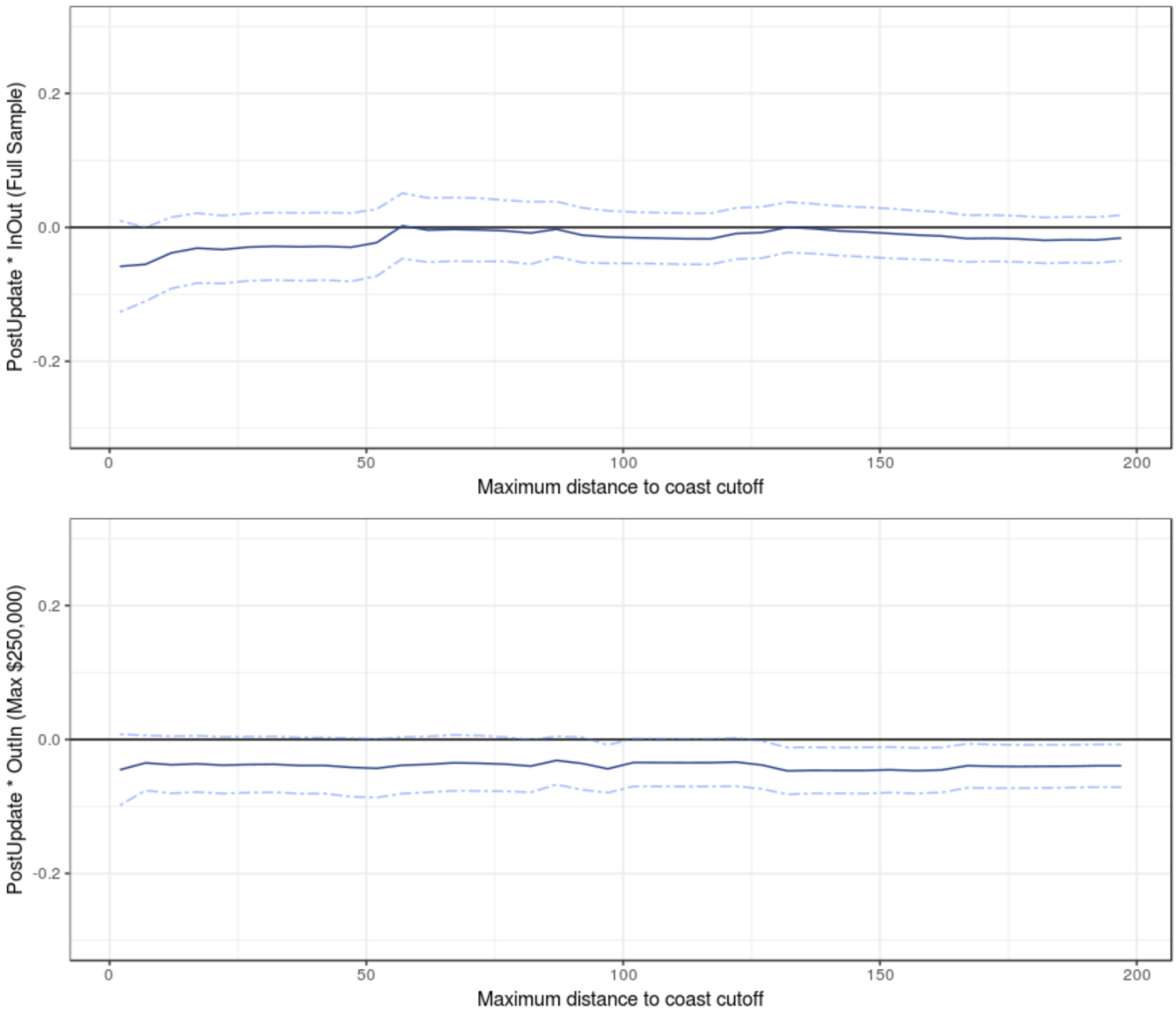
	Dependent variable:			
	Ln Price			
	(1)	(2)	(3)	(4)
Post Update	0.205***	0.023***	0.205***	0.023***
	(0.007)	(0.008)	(0.007)	(0.008)
High Awareness Month			0.008	−0.009
			(0.006)	(0.006)
Post Update * OutIn	0.018	−0.017	−0.001	−0.034
	(0.024)	(0.026)	(0.024)	(0.026)
Post Update * OutIn * High Awareness DMA	−0.109*	−0.071		
	(0.057)	(0.049)		
Post Update * OutIn * High Awareness Month			0.0002	−0.006
			(0.054)	(0.050)
Sample	OutOut, OutIn	OutOut, OutIn	InIn, InOut	InIn, InOut
Minimum Transaction Price		250,000		250,000
Ref cat	OutOut	OutOut	InIn	InIn
Year dummy	Y	Y	Y	Y
Month dummy	Y	Y	Y	Y
Property FE	Y	Y	Y	Y
R2	0.872	0.893	0.872	0.893
Within R2	0.065	0.062	0.065	0.062
Observations	39,842	9,582	39,842	9,582

*p<0.1; **p<0.05; ***p<0.01

This table presents the results of the unit fixed effect model presented in equation (3) augmented with measures of awareness on the panel dataset of repeated sales restricted to *OutOut* and *OutIn* properties. Real estate transaction data and house characteristics are from Zillow. Floodplain data are from FEMA. I trim the real estate dataset to exclude the bottom and top 1% sales amount. Properties sold in counties mentioned in a PDD between 2006 and the month of the first sale are excluded from the sample. In columns (1) and (2), *High Awareness DMA* is a dummy equal to 1 if average Google Index in the Designated Market Area (DMA) where is located the property is above the median of Google Index averages. In column (3) and (4), *High Awareness Month* is a dummy equal to 1 if the monthly-DMA Google Index is above the 75th percentile of the sample's monthly-DMA indexes. All specifications include property fixed effects as well as year and month dummies. Unlike odd columns, results presented in even columns are estimated on a subset of the sample excluding properties sold for less than $250,000. Standard errors are presented below estimates and are not clustered.

Robustness to coastal sampling

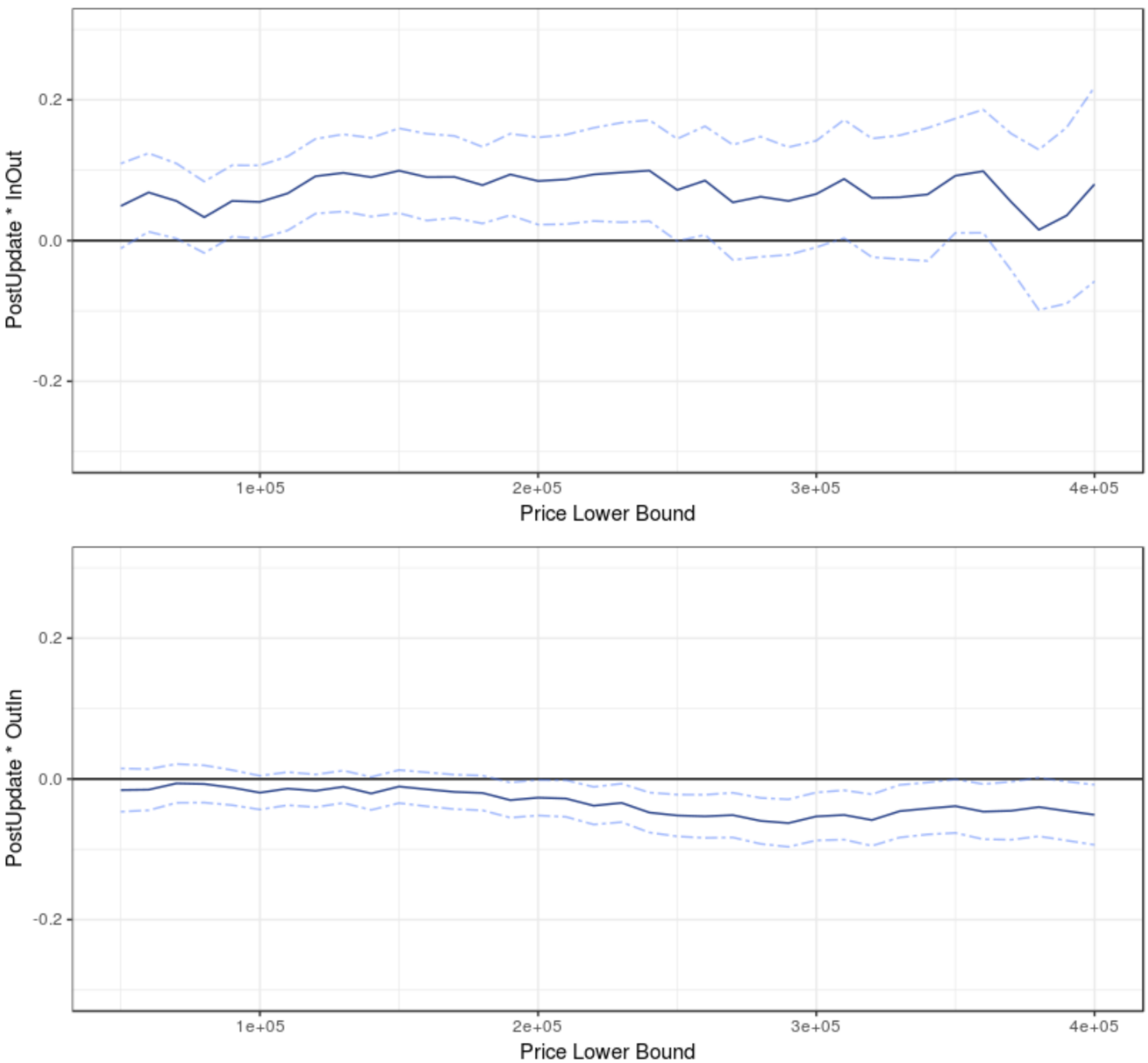
Figure 12: Risk Reclassification Impact - Different Distance-to-coast thresholds



The graph on the top of the figure displays the evolution of the *Post Update * OutIn* coefficient for the risk-reclassification unit fixed effect model when moving up the distance-to-coast threshold used to define the sample. The specification is similar to the one presented in column (1) of Table 11.
The graph on the bottom of the figure displays the evolution of the *Post Update OutIn* coefficient for the risk-reclassification unit fixed effect model when moving up the distance-to-coast threshold used to define the sample. The sample is restricted to properties with a transaction price above $250,000. The specification is similar to the one presented in column (2) of Table 11. 90% confidence intervals are represented by the dotted blue lines.

Robustness to Transaction Price Lower Bound

Figure 13: Risk Reclassification Impact - Different Price Lower Bounds



The graph on the top of the figure displays the evolution of the *Post Update * InOut* and coefficient for the risk-reclassification unit fixed effect model when moving up the lower bound of the transaction price of properties included in the sample. Specifications are similar to those presented in columns (3) and (4) of Table 11. The reference category are *InIn* properties sold before the map update.
The graph on the bottom of the figure displays the evolution of the *Post Update * OutIn* coefficient for the risk-reclassification unit fixed effect model when moving up the lower bound of the transaction price of properties included in the sample. Specifications are similar to those presented in columns (1) and (2) of Table 11. The reference category are *OutOut* properties sold before the map update. 90% confidence intervals are represented by the dotted blue lines.

Robustness to Different Awareness Definitions

Table 23: Event Study: The role of awareness during the updating process Different Awareness Definitions

	Dependent variable:	
	Ln Price	
	(1)	(2)
Pre Update Process * High Awareness (1b)	−0.001 (0.032)	
Pre Update Process * High Awareness (2b)		−0.012 (0.028)
Preliminary BFE	−0.045 (0.087)	−0.072 (0.086)
Preliminary BFE * High Awareness (1b)	−0.107 (0.156)	
Preliminary BFE * High Awareness (2b)		-0.414^{***} (0.089)
Appeal Period	0.029 (0.161)	0.026 (0.108)
Appeal Period * High Awareness (1b)	−0.045 (0.210)	
Appeal Period * High Awareness (2b)		−0.122 (0.162)
Map Pending	−0.041 (0.085)	−0.090 (0.090)
Map Pending * High Awareness (1b)	0.012 (0.101)	
Map Pending * High Awareness (2b)		0.099 (0.128)
Map Update Month	−0.021 (0.064)	−0.022 (0.062)
Map Update Month * High Awareness (1b)	0.011 (0.028)	
Map Update Month * High Awareness (2b)		0.017 (0.031)
Post Update	0.099 (0.299)	0.056 (0.121)
Post Update * High Awareness (1b)	−0.334 (0.305)	
Post Update * High Awareness (2b)		−0.062 (0.401)
Sample	OutIn	OutIn
Minimum Transaction Price	250,000	250,000
Size Controls	Y	Y
T*B*C FE	Y	Y
Y*B*C FE	Y	Y
R2	0.756	0.756
Within R2	0.062	0.062
Observations	5,984	5,984

*p<0.1; **p<0.05; ***p<0.01

This table presents the results of the event study on the role of awareness during the flood map updating process (see figure 6) for 3 different definitions of *High Awareness*. The sample is restricted to properties experiencing risk reclassification after their community's map update (*OutIn* properties) and valued above $250,000. All columns have the same specification than columns (3) and (4) of Table 13. In column (1), *High Awareness (1b)* is a dummy equal to 1 if Google Index (1) is greater or equal to the mean of the sample's monthly-DMA indexes. In column (2), *High Awareness (2b)* os a dummy equal to 1 if Google Index (2) is greater or equal to the mean of the sample's monthly-DMA indexes. Google Index (1) and Google Index (2) are defined in Table 13. Estimates are consistent across the two awareness definitions and significant for *Preliminary BFE * High Awareness* for the second awareness definition.