Salience and Households' Flood Insurance Decisions

Zhongchen Hu*

London School of Economics

December 5, 2020

Abstract

Flooding is one of the most costly natural disasters faced by US households, yet policymakers are puzzled by the low take-up rates for flood insurance. In this paper, I argue that households' insurance purchases are affected by the low salience of flood risk. Leveraging novel transaction-level data, I use two empirical strategies to support my hypothesis. First, I exploit a staggered national reform that publicizes already freely-available flood risk information across US counties. Insurance purchases, in response to the reform, increase by 30.6%, with the strongest effect in counties where the ex-ante salience of flood risk is low. Second, I exploit variation in salience induced by flood events shared through social media. Households purchase significantly more flood insurance after their geographically distant peers experience floods. My results suggest that behavioral frictions have a major impact on households' insurance decisions.

*I am thankful for the advice of Ashwini Agrawal, Fabrizio Core, Juanita González-Uribe, Isaac Hacamo, Dirk Jenter, Peter Koudijs, Dong Lou, Daniel Paravisini, Cameron Peng, and Su Wang. I am also grateful to seminar participants at the London School of Economics for helpful comments. Address for correspondence: z.hu5@lse.ac.uk

1 Introduction

Flooding is one of the most costly natural disasters in the US and is drawing increasing attention in public policy debates. In the 2010s, flood-related disasters have caused losses of $658 billion.[1] As flood risk materializes with low frequency but has catastrophic consequences for households' asset values and welfare, insurance is crucial for households to hedge this tail risk. However, the take-up rate of even subsidized flood insurance is low in the US.[2] Policymakers are looking to encourage insurance purchases as the expected cost of uninsurance is high.[3]

Yet why would flood insurance be underutilized by households? I test the hypothesis that households do not buy flood insurance because the risk is not fully salient to them. I define salience as the phenomenon when one's attention is disproportionally directed to one portion of the environment and the information it contains (Taylor and Thompson, 1982). Models of salience and heuristics (Tversky and Kahneman, 1973, 1974; Bordalo et al., 2012, 2013; Kőszegi and Szeidl, 2013) suggest that the low salience of flood risk (in normal times) might lead to inattention, underestimation, and low insurance take-up.

Identifying a causal effect of salience on insurance demand requires overcoming several challenges. First, the salience of flood risk is difficult to measure, as we do not observe households' attention. Second, the determinants of insurance demand—such as information about flood risk, actual risk, and attention—often move simultaneously, making it difficult to isolate the key driver. For example, in the aftermath of a flood, salience likely increases, but so may the actual flood risk, due to damage to infrastructure or to changes in geological conditions. Third, as the equilibrium price and quantity of insurance transactions are jointly determined by supply and demand, it is difficult to distinguish shifts in the supply curve from shifts in demand.

I overcome these challenges by using data from the National Flood Insurance Program (NFIP) and two quasi-experiments. The NFIP was created by the US Congress in 1968 to provide affordable household flood insurance. I obtain over 50 million transaction-level

[1] The estimate is obtained from the National Oceanic and Atmospheric Administration.

[2] According to the Insurance Information Institute 2016 Survey, only 12% of US households had flood insurance. Consistently, only 17% of the flooded homes by Hurricane Harvey in 2017 were insured. Similar concerns have been documented for other insurance products and in other countries (Cole et al., 2013; Karlan et al., 2014; Banerjee et al., 2019; Finkelstein et al., 2019).

[3] According to the National Flood Insurance Program, the cumulative difference between premiums and claims is −$5.85 billion over the past 20 years, implying that, in aggregate, policyholders would have accumulated an additional loss of $5.85 billion without flood insurance.

observations from January 2009 to August 2019, including information on policy start and end dates, premiums, coverage, house characteristics, and locations.

The NFIP is ideal for measuring shifts in insurance demand for two reasons. First, the supply is perfectly elastic. The insurance rates are fixed conditionally on given risk profiles, which primarily depend on government-designated risk zones. They do not otherwise vary by state, locality, or market conditions. Second, unlike most other property and casualty risks, flood risk has been shunned by private insurers (Horn and Webel, 2019), leaving few outside options for households.[4] Thus, standard models predict that the marginal household will go from not buying to buying flood insurance from the NFIP when its hedging demand becomes stronger. Aggregating at the county-month level, the change in the number of policies in-force thus captures shifts in the demand curve.

To identify the effect of salience, my first quasi-experiment exploits the NFIP's staggered transformation of existing paper-based flood risk maps into digital ones. Figure 1 presents an example. This reform increased flood risk salience via local marketing campaigns (such as open houses and newspaper advertisements) to publicize the digital maps (see Section 2.2 for details). In addition, the new maps are more eye-catching, highlighting flood risk with visually appealing colors and satellite street views. Enhanced visual salience has been shown to affect human behavior in contexts such as stock investments (Towal et al., 2013; Li and Camerer, 2019; Bose et al., 2020).

Under the null hypothesis of perfect attention to flood risk, this reform should not significantly affect households' insurance purchases, for two reasons. First, before digitalization, the risk information was already freely obtainable, as the scanned maps were always available online. Second, the hard information about flood risk is identical in the two maps in Figure 1.

Using a difference-in-differences strategy that exploits the county-by-county introduction of digital flood risk maps, I show that the number of flood insurance policies in-force increases by 30.6% after the new maps are published. I find no evidence of a differential pre-trend in the treated counties. Following digitalization, the increase in insurance demand materializes quickly over several months and appears to be permanent, as I find that the incremental policies are renewed in subsequent years. This result is

[4] The private market for flood insurance barely exists because the government heavily subsidizes the NFIP (see footnote 8). Cutting subsidies and privatization are the subject of contemporary policy debates, but irrelevant to this paper's research question.

consistent with the hypothesis that households are inattentive to flood risk due to its low salience. My quasi-experiment cannot distinguish whether the risk information was known by households but ignored or whether it was simply unknown. Both are examples of limited attention to flood risk, as the information cost in my setting is minimal.[5]

I provide a set of additional empirical findings consistent with the salience hypothesis. First, I show that average house prices (measured by Zillow's home value index) in the treated counties drop by 2% after digitalization, suggesting that when flood risk becomes more salient, it is more heavily priced into real estate values. This finding is consistent with households underweighting product attributes (such as a property's exposure to flooding) until they are made salient (Bordalo et al., 2013).

Second, I show that the digitalization effect is *weaker* where the salience of flood risk is already high. Specifically, flood occurrences in the past predict a smaller post-digitalization increase in flood insurance purchases.[6] This correlation is monotonic—the longer the county has not had a flood, the larger the digitalization effect. In a similar vein, I also show that, given any value of n from 1 to 6, counties with no flood in the past n years before digitalization experience a larger treatment effect than do counties that had at least one flood in that period.

Third, I consider several proxies for households' pre-digitalization attention to flood risk. I find that the digitalization effect is *weaker* in counties with greater incomes, higher education, more believers in global warming, or higher flood risk. The strongest effect is observed in moderate-risk areas, likely because flood risk is already salient in high-risk areas and because flood insurance is not needed in low-risk areas.

I consider several alternative explanations for my findings. First, the relationship between digitalization and flood insurance demand might not be causal if both are driven by a third factor. One alternative explanation is that the digitalized maps are introduced across geographies that experience increasing underlying flood risk, which simultaneously drives up insurance purchases. In other words, the digitalization process might endogenously capture an existing trend in underlying flood risk and insurance demand. However,

[5]Even before digitalization, googling "flood insurance" or "flood risk" would instantly bring up information about the NFIP and its flood hazard maps (scanned copies). Thus, ignorance of the importance of hedging flood risk is effectively equivalent to ignorance of flood risk information.

[6]There is a large literature showing that past experiences affect subsequent decision making. See, for example, Malmendier et al. (2011); Malmendier and Nagel (2011); Dittmar and Duchin (2016); Malmendier and Nagel (2016); Bernile et al. (2017); Schoar and Zuo (2017); Kuchler and Zafar (2019).

the parallel pre-trend, shown in Figure 5, refutes this explanation.

A related non-causal explanation is that the publication of digital maps follows a local flooding event, which simultaneously induces insurance purchases (Gallagher, 2014). This explanation differs from the previous one if flooding is unpredictable, in which case the pre-trend test does not help. Therefore, I provide additional evidence to cast doubt on this alternative explanation. First, past floods cannot predict digitalization timing. Second, there is no change in flood frequency around digitalization, while the alternative hypothesis predicts bunching before it. Third, anecdotally, local officials explicitly declared that the publication of new digital maps is unrelated to recent floods.[7]

Next, I investigate whether there is a purely rational explanation for why the reform leads to increasing demand for flood insurance. While my data does not allow me to fully rule out all rational hypotheses discussed below, I present a series of evidence suggesting that my findings are unlikely to be driven by purely rational mechanisms.

The first purely rational hypothesis I consider is that having digital maps on the Internet reduces households' transaction cost of finding flood risk information. Some previously uninformed and uninsured households may see the maps online and start buying insurance. However, this is unlikely, as flood risk information was already freely available to households even before digitalization. In particular, the scanned maps (as in Figure 1) were always online, and people could also easily acquire information from local NFIP agents or toll-free hotlines. I provide further evidence to support this argument in Section 4.6.2. For example, I find that the reform also affects the intensive margin of existing policyholders, who are already informed by their current policies (the most crucial piece of information being their risk zones). Yet, they increase their insurance coverage, which is unlikely due to a reduction in information cost.

The second purely rational hypothesis I consider is that the informational content of the paper and digital maps is fundamentally different. Anecdotally, some areas of some maps might have been updated, while being digitalized, to reflect higher or lower flood risk. However, I present evidence in Section 4.6.3 suggesting that the scope of updating is likely small, consistent with the examples in Figure 1 and the Web Appendix. For example, I show that digitalization does not cause existing policyholders to cancel or to

[7] For example, on the official website of the City of Alexandria, Virginia, it says, "This effort [developing new digital maps] is unrelated to recent flooding the city has experienced from flash flooding in July 2019 and, more recently, on July 23 [2020]." See Section 4.6.1 for details.

stop renewing their flood insurance policies. If the map modification had been extensive, we should observe cancellations induced by risk downgrades.

To further support the salience hypothesis and avoid the information-based alternative explanations, my second strategy exploits variation in social connectedness on Facebook and leverages non-local flooding events. This strategy complements the first one by making it unequivocal that the local flood risk does not change and that there is no new information about the local risk. Specifically, for a given flooding event (e.g., in Boston), I examine flood insurance purchases in geographically distant states. Within the same far-away state (e.g., California), I compare changes in flood insurance purchases in counties that are more versus less socially connected to Boston, before and after the flood in Boston. The idea is that when an individual sees Facebook friends sharing flood experiences, the possibility of going through a flood becomes more salient.

Pooling all major floods (that triggered federal assistance) between 2010 and 2019 in an event study design, I find that the number of flood insurance policies in-force increases by 1%–5% in counties that are more socially connected to the flooded area, compared to the less connected counterfactuals in the same distant state. I find no pre-trend, and I document two additional findings consistent with the salience hypothesis. First, the effect is monotonic in the strength of social connectedness. Second, the most damaging floods cause the most pronounced effect across social networks. My strategy is similar to that of Bailey et al. (2018a), who show that friends' house-price experiences affect one's own housing investment decisions. As regional housing markets are possibly correlated while geographically distant floods are much less likely to be, my setting makes it arguably easier to rule out rational learning as an alternative explanation.

This paper does not claim that the increase in flood insurance purchases—or that the NFIP in general—is socially optimal. It is likely welfare-improving for households, as the NFIP offers heavily subsidized insurance rates, and it is the government's declared goal to encourage more people to acquire flood insurance.[8] A back-of-the-envelope calculation suggests that the expected net benefit of buying flood insurance is $1,240 per year per flood-prone household.[9] However, a complete welfare analysis needs to take the subsidies

[8]The NFIP suggests that eliminating the subsidy would cause aggregate premiums to increase by 50-75% (Hayes and Neal, 2011). Consistently, the NFIP's premiums cannot fully cover claims (see footnote 3), and its operating expenses further worsen its financial condition—it owed $20.5 billion to the Treasury as of December 2019 (excluding a $16 billion debt canceled by Congress in 2017).

[9]The estimation is based on the following assumptions: an NFIP-defined 1% inundation probability

and government expenditures into account, which is beyond the scope of this paper.[10]

This paper contributes to a rapidly growing literature on behavioral household finance and is among the first to study households' insurance decisions against rare disaster risks. There is considerable evidence that households make suboptimal decisions because of limited attention. For recent empirical findings from various contexts, see Barber and Odean (2008); Chetty et al. (2009); Choi et al. (2009); Finkelstein (2009); Brown et al. (2010); Malmendier and Lee (2011); Lacetera et al. (2012); Hastings and Shapiro (2013); Stango and Zinman (2014); Busse et al. (2015); Andersen et al. (2020). Acquiring insurance against low-probability disasters is a critical household decision, but it is understudied by the literature. My results suggest that households are especially vulnerable to behavioral biases when assessing low-salience tail risks. This finding is important to policymakers, as neglecting rare disasters can be costly for households.

A closely related paper is Gallagher (2014), which shows that flood insurance purchases increase after a local flooding event. My work differs in several dimensions. First, Gallagher (2014) documents a long-term but declining effect of personal experiences, while I provide evidence of a persistent effect in a different setting.[11] Second, his NFIP data is yearly, aggregated, and covers an earlier period (1980-2007). Third, his results are open to different interpretations: actual risk may change in the flooded area (e.g., due to foundation erosion), households may learn about flood risk and exposures from local and nearby floods, and households may overweight recent experiences.

This paper also adds to the nascent literature on the effects of climate risk on household behaviors and economic outcomes. A number of studies examine whether climate risk, in particular sea-level rise, is capitalized into real estate values (Giglio et al., 2015; Keenan et al., 2018; Bernstein et al., 2019; Baldauf et al., 2020; Murfin and Spiegel, 2020) and mortgages (Issler et al., 2019; Ouazad and Kahn, 2019). Other papers study how personal experiences of climate change or natural disasters affect beliefs about climate risk (Li et al., 2011; Zaval et al., 2014; Dessaint and Matray, 2017; Chang et al., 2018; Anderson and Robinson, 2019; Choi et al., 2020). These studies typically demonstrate a short-term impact of a shock (e.g., a day of unusual weather). My focus on public

p.a., an average premium of $993, an average coverage of $227,112, and an average deductible of $3,831.

[10] Wagner (2019) provides a framework for studying the welfare effects of the NFIP. Her results suggest that enforcing a flood insurance mandate will increase social welfare.

[11] In Gallagher (2014), the effect is rather stable around 8% in event years 1 to 8, but it substantially drops in years 9 and 10. In contrast, Appendix Figure A.4 shows a persistent effect in my setting.

awareness and visual salience is novel, and the substantial effects of campaigns and social networks that I document have unique implications for effective climate policy.

The remainder of the paper is organized as follows. Section 2 describes the institutional background and details the data. Section 3 describes the two empirical strategies. Section 4 presents findings from the staggered introduction of digital flood risk maps. Section 5 presents results using social media connectedness and geographically distant floods. Section 6 concludes.

2 Institutional Background and Data

2.1 The National Flood Insurance Program

The US Congress founded the National Flood Insurance Program (NFIP) in 1968, and as of August 2019, it covers all 50 states and 3,053 (out of 3,143) counties. The program creates flood hazard maps for participating communities (subdivisions of counties, such as townships, villages, and cities), and only residents in participating communities are eligible to buy NFIP policies. The insurance premiums primarily depend on risk zones (set centrally by the government) and also vary by the choices of coverage and house characteristics. They do not otherwise vary by state, locality, or market conditions.

The NFIP data is maintained by the Federal Emergency Management Agency (FEMA). I obtain more than 50 million transaction-level observations between January 2009 and August 2019, including policy effective dates, policy termination dates, premiums, coverage, deductibles, first policy dates, cancellation dates, and house characteristics. The policies are annually renewed, and renewals appear as separate transactions in the dataset. Broad location information (such as census tract, county FIPS code, and community code) is available, but specific properties cannot be identified due to privacy protection (geographic coordinates are truncated to one decimal point).

Based on the policy effective dates and termination dates, I can calculate the number of policies in-force (a stock measure) and the number of policies purchased (a flow measure) in a given month for a given county. By definition, the change in the number of policies in-force between two consecutive months $t-1$ and t is equal to the number of policies purchased minus the number of policies expired in t. Additionally, knowing the policyholder's first policy date allows me to determine whether the anonymized transaction is a first-time purchase or a renewal. Since I do not observe data for 2008, I can not perfectly impute the number of policies in-force in 2009. Therefore, my analysis starts from January 2010. The data granularity and the 10-year panel allow me to zoom in on households' flood insurance demand in the very short-run as well as to keep track of its long-term dynamics.

Table 1 presents descriptive statistics of the NFIP. Panel A shows that in an average month, there are 5.29 million policies in-force nationally. From these policyholders, the program receives \$3.32 billion in premium for \$1.26 trillion in coverage.[12] The nationwide average annual premium per policy is \$628, and the average coverage per policy is \$238 thousand. Panel B shows that the cross-sectional heterogeneity is stark. While the average county purchases 1,766 policies, the median is only 120. Appendix Figure A.1 shows a geographical heat map of the number of policies in-force. As one would expect, coastal counties have the highest densities. The variation in the average insurance premium, reported in Panel B, is due to differences in flood risk across counties, rather than to price differentiation.

2.2 Digitalizing Flood Hazard Maps

The Map Modernization (Map Mod) program was launched in 2003 by FEMA to transform the flood hazard maps from paper to digital. These improvements have continued under FEMA's successor Risk Mapping, Assessment, and Planning (Risk MAP) program launched in 2009. Before digitalization, people could obtain paper maps from local NFIP agents or offices, find scanned copies online, or call toll-free NFIP hotlines to acquire relevant information (e.g., risk zone designations).

Figure 1 shows an example—the Town of Colfax (community code: 220077), Grant Parish (county code: 22043), Louisiana—to illustrate what users can see on the flood hazard maps (available at msc.fema.gov/portal). Figure 1.a shows a scanned copy of the legacy paper map. It was published on November 16, 1995 and was in use until the digital map (Figure 1.b) became available on June 16, 2016. These two maps convey identical information about the flood risk in the Town of Colfax (although the jurisdiction boundary is slightly different). The blue areas in Figure 1.b are the Special Flood Hazard

[12] Compared to other property and casualty insurance in terms of aggregate premiums (2017 data): earthquake (\$2.9B), aircraft (\$1.5B), mortgage guaranty (\$5.0B), burglary (\$0.3B), and fire (\$11.6B).

Areas (SFHA) which are expected to have a 1% (or higher) flooding probability per year (referred to as 100-year floodplains), and the brown areas are of median risk, with a 0.2% (to 0.99%) flooding probability per year (referred to as 500-year floodplains).

2.2.1 Local Campaigns

Around the publication of the new digital maps in a county, FEMA instructs the local government to run a campaign (known as the "Flood Risk Open House") to increase public attention. FEMA provides customized marketing packages and templates (known as the "Toolkit") to advertise the Open House by placing advertisements in local newspapers and on radio, distributing flyers, and posting announcements on community websites and social media. The full toolkit can be found on FEMA's website.[13] Figure 2 presents examples of actual advertisements and announcements.

2.2.2 Visual Salience

Visually, the new digital maps are more appealing and might increase the salience of flood risk, even if the hard information about flood risk does not change on the maps. Figure 1.b appears visually more salient than Figure 1.a because flood risk is more intuitively highlighted by the water-like color and because the visibility of actual houses and streets in satellite views might create resonance and draw attention.

To assess which format is more eye-catching in a more systematic manner, I leverage a machine-learning-based methodology called the Saliency Attentive Model (SAM), developed by Cornia et al. (2018). The model predicts human eye fixations on an image within the first few seconds of gaze. I put the digital and paper maps side by side as one input image. Appendix Figure A.2 presents the output returned by SAM; the overlay heat map represents the predicted human eye fixations. The results show that the colored digital map draws more attention than the black-and-white paper map.

2.3 Staggered Introduction of Digital Maps

Revisiting the digital map in Figure 1.b, the dates shown on each small area indicate the publication dates of the flood hazard maps. For instance, the Town of Colfax published its digital map on June 16, 2016, whereas the neighboring county to the west did so on

[13] See http://townofvanburen.com/wp-content/uploads/2016/08/Onondaga-Open-House-Community-Packet-FINAL.pdf, for an example of a tailored toolkit that FEMA sent to Onondaga County, NY.

July 6, 2015. For the dot-shaded area to the south-west (part of a different county), its digital map is not yet available, and the latest version is still the paper map from September 5, 1984.

I obtain all communities' map publication dates from FEMA's Community Status Information (CSI) database. The data suggests that the digitalization programs are run at the county level, and communities within the same county typically publish their new maps simultaneously, consistent with the open house examples discussed in the previous section. Since the other essential data for my analysis, such as flood occurrences and housing prices, is not as granular as communities, I examine the staggered introduction of digital maps at the county level. When there are disparities among communities, I define the treatment time of a county as the calendar month in which more than 50 percent of its communities simultaneously publish their digital maps.[14] I verify that the results are robust to various alternative definitions of county-level rollout (see Section 4.1 for details).

Figure 3 maps the staggered introductions of digital flood hazard maps by time and county. The darker the shade, the more recent the treatment dates; and the unshaded areas are the untreated. There is no discernible clustering, and there is obvious variation within regions and states. (Figure 1.b is a consistent micro example showing three adjacent counties with different publication timing.) I also verify that the digitalization timing is uncorrelated with either the level of flood risk (see Appendix Figure A.3) or past flood occurrences (see Section 4.6.2).

One limitation of the CSI database is that it only records the latest map publication dates, while FEMA aims to review its maps every five years. Thus, a concern is that some counties may have published digital maps twice, but I only observe the latter. This might lead to underestimation, as the first treatment is ignored and the county is coded as control until the second treatment.[15] However, Appendix Section A presents evidence suggesting that FEMA largely falls short of its goal. Hence, the date I observe is likely to be the county's only map publication in the 2010s.

[14] The 50 percent threshold is not a cumulative measure. Instead, it means at least half of the communities publish the digitalized maps simultaneously in one specific month. This criterion by construction captures the unique treatment event at the county level.

[15] Presumably the first treatment is stronger as it combines a marketing campaign and an enhancement to visual salience (from paper to digital), whereas the second treatment has only the campaign.

2.4 Presidential Disaster Declaration

I identify flooding incidents using the Presidential Disaster Declaration database. The declaration process was established in 1988 (by the Stafford Act) for local and state governments to request federal natural disaster assistance. This database provides information on disaster ID numbers, declaration dates, incident begin and end dates, declared states and counties, and incident types. I categorize certain incident types—Severe Storm, Hurricane, Flood, Coastal Storm, and Typhoon—as flood-related events. I identify 419 flood-related declarations over my sample period; one declaration typically includes multiple affected counties.

2.5 Social Connectedness by Facebook

Bailey et al. (2018b) aggregate anonymized information from the universe of friendship links between all Facebook users as of April 2016 to produce a county-by-county social connectedness measure. I initially obtained the data through a non-disclosure data-sharing agreement; the data was later made open source by Facebook. Bailey et al. (2018b) calculate the Social Connectedness Index (SCI) for a pair of counties as the number of Facebook friendship links between individuals located in those two counties. They further create a measure called the *relative probability of friendship* by dividing the SCI for county i and j by the product of the number of Facebook users in the two counties. If this measure is twice as large, it means that a given Facebook user in county i is about twice as likely to be connected with a given Facebook user in county j. I denote the *relative probability of friendship* by $p_{i,j}$ and use it to measure county-by-county social connectedness.

3 Empirical Strategies

3.1 Staggered Risk Maps Digitalization

My first strategy exploits the staggered introduction of digital flood hazard maps, which increase the salience of flood risk by enhancing public awareness via campaigns and by presenting risk information in a visually more salient format. Leveraging variation in digitalization timings across counties, the identifying assumption of my difference-in-differences strategy is that, without digitalization, outcomes would have moved in parallel in the treated and untreated counties.

I first estimate the following canonical staggered difference-in-differences regression, which models treatment events as immediate and permanent shifts in the outcome:

$$Y_{it} = \alpha_i + \lambda_t + \beta * Digitalization_{it} + \epsilon_{it} \tag{1}$$

In this two-way fixed effects model, unit and time are specified as county and year-month. Y_{it} measures the number of policies in-force in county i at time t. I normalize the number in January 2010 to 100 for each county, so that the results are not dominated by extremely large counties. Let t_i^* denote the digitalization time for county i. $Digitalization_{it}$ is an indicator variable $\mathbb{1}(t > t_i^*)$ that turns on if county i has published its new digital flood hazard map at time t; this term is set to zero for untreated counties for any t. α_i and λ_t are unit and time fixed effects, respectively. The coefficient β measures the change in the outcome following digitalization. Standard errors are clustered at the county level to allow for arbitrary dependence of ϵ_{it} across t within i.

The identifying assumption requires the digitalization timing to be uncorrelated with the outcome. If this assumption is not satisfied, the treated counties might already diverge from the controls before the treatment date. Also, the change in flood risk salience, due to the digitalization, might not affect households' insurance decisions immediately; instead, the impact might develop gradually. To accommodate these possibilities, I also estimate the following nonparametric model:

$$Y_{it} = \alpha_i + \lambda_t + \sum_{k=-\underline{L}}^{k=\overline{L}} \beta_k * \mathbb{1}(t = t_i^* + k) + \epsilon_{it} \tag{2}$$

The primary advantage of regression (2) is that it allows me to flexibly and visually assess the pattern of outcomes relative to the event date. $\{\beta_k\}$ for $k < 0$ correspond to the pre-trends, and $\{\beta_k\}$ for $k > 0$ measure the dynamic effects of the digitalized maps. These effects are measured relative to β_{-1}, which is omitted.

3.2 Social Connectedness and Geographically Distant Floods

My second strategy exploits exogenous variation in flood risk salience caused by cross-sectional variation in social connectedness with a geographically distant flooded area.

The idea is that when an individual sees Facebook friends sharing flood experiences, the possibility of going through a flood becomes more salient.

For a given flooding event f, I identify a set of flooded counties $\{j\}_f$ and a set of geographically distant states and counties $\{i\}_f$ (that are at least 750 miles away from the flooded area). Then, within a far-away state, I define the treatment (control) group as counties that are more (less) socially connected with the flooded area. As the flooded area $\{j\}_f$ may consist of several counties, I calculate county i's social connectedness with the flooded area as a weighted average of the county-by-county relative probability of friendship $p_{i,j}$ (discussed in Section 2.5):

$$p_{i,f} = \sum_{\{j\}_f} w_j * p_{i,j} \tag{3}$$

where w_j represents population-weighting or equal-weighting. Within a *state*, county i is coded as treated (control) with respect to event f, if $p_{i,f}$ is above (below) its state median.

Figure 4 depicts one specific flooding event—Hurricane Florence in September 2018. It shows a heat map of social connectedness ($p_{i,f}$, blue-shaded) with the flooded area (red-shaded). The uncolored counties located less than 750 miles away from the flooded area are not included in this particular event study with respect to Hurricane Florence.

I stack individual event studies (with respect to every flooding event f), and the standard errors are clustered at the county level. Each event study features the following difference-in-differences regression:

$$Y_{it} = \beta_0 + \beta_1 * Connected_i \times Post_t + \beta_2 * Connected_i + \beta_3 * Post_t + \epsilon_{it} \tag{4}$$

$Connected_i$ is the treatment dummy (that turns on if i is eventually treated); $Post_t$ is the post-event dummy (that turns on if t is after the flooding time). The outcome variable Y_{it} measures the insurance demand in county i at time t, which is defined as in regression (1). The interaction $Connected_i \times Post_t$ is the key explanatory variable of interest.

By construction, I define treated and control counties conditionally on being in the same state, so that they are likely to have similar climatological and economic conditions. Even if counties in the same state have distinct climates and flood risk, the difference-in-differences construct of regression (4) teases out the fixed differences.

More formally, in order to interpret the estimate of β_1 as the causal effect of friends' flood experiences, I need to assume that insurance purchases in treated and control counties would have evolved in parallel without treatment (i.e., a distant flooding event). I test for parallel pre-trends by replacing $Post_t$ with a sequence of event time dummies $\{\mathbb{1}(t = t^* + k)\}$. The coefficients $\{\beta_1^k\}$ on the interaction terms $\{Connected_i \times \mathbb{1}(t = t^* + k)\}$ allow me to examine the patterns in insurance purchases in the months before and after the flood experienced by geographically distant friends. I present evidence of parallel pre-trends in Section 5.1. I discuss potential concerns about this event study approach in Section 5.3 and consider an alternative empirical methodology.

4 Empirical Findings: Digitalization of Risk Maps

In the following two sections, I test the hypothesis that households will demand more flood insurance if the salience of flood risk increases. Section 4 presents the results of my first quasi-experiment, which exploits the staggered introduction of digital flood hazard maps across US counties. I also provide additional evidence supporting the hypothesized channel of salience and evaluate several alternative explanations.

4.1 Main Results: Digitalization and Insurance Purchases

Table 2 presents the staggered difference-in-differences estimates of regression (1). I examine the change in the number of policies in-force (with January 2010 normalized to 100). The standard errors are clustered at the county level to adjust for autocorrelation.[16] Column 1 shows that the average number of insurance policies goes up dramatically by 21.39, following the digitalization of flood risk maps. Note that this is the mean treatment effect over the entire post period, and it corresponds to an increase of 19.91% relative to the mean of 107.42 at event time zero.

The result is robust to a variety of alternative specifications and sample restrictions. For example, in column 2, I add county-level covariates, including the average premium per policy and the average coverage per policy. In column 3, I exclude the never-treated counties from my analysis; thus, to control for any underlying trends, the staggered

[16]In fact, the clustered standard errors are almost 10 times larger than the non-clustered ones, suggesting strong positive autocorrelation. This is not surprising, as the policies are renewed annually.

difference-in-differences estimation uses only counties that have not yet had digitalization or have already had digitalization. This specification addresses the concern that some never-treated counties (e.g., located in the desert) are poor counterfactuals. In both columns, the coefficient estimate has a similar magnitude and statistical significance.

I assess the robustness of my treatment construction in several dimensions. First of all, column 4 of Table 2 considers a continuous treatment approach and obtains a similar result. As detailed in Section 2.3, in my baseline discrete specification, a county is defined as treated if more than half of its communities publish digital maps in the same month. In column 4, I refine the treatment indicator $Digitalization_{it}$ to be the cumulative fraction of digitalized communities. In the extreme case, in which all communities always publish digital maps simultaneously, the continuous and binary approaches are identical.

Appendix Table A.1 addresses the concern that population is not evenly distributed across communities. I refine the treatment definition to be population-weighted; that is, a county becomes treated in a given month if more than half of its population gets access to digital flood maps in that month. Since population data is not available at the community level, I use the number of policyholders as a proxy. Columns 1 through 3 show that my results hold. Additionally, in columns 4 and 5, I use different thresholds to construct the binary treatment; the results are consistent.

Appendix Table A.2 shows that the results are also insensitive to sample selection. In my baseline specification, I estimate regression (1) using the full sample and do not impose any restrictions on leads and lags. In this table, I restrict the event window by various specifications of different leads and lags. The estimates demonstrate a similar magnitude and statistical significance across all specifications.

Figure 5 shows the dynamic effects of digitalization on insurance purchases. It plots the coefficient estimates of the event time dummies from regression (2). Crucially, there is no evidence of a differential pre-trend in the treated counties. This addresses the concern that the treatment assignment endogenously captures trends in the outcome variable; for instance, the government might expedite digitalization in counties that are experiencing an increase in underlying flood risk. The parallel pre-trend rejects such an explanation and validates the identifying assumption (see more discussions in Section 4.6.1).

Figure 5 also shows that, following digitalization, insurance purchases rise sharply. The impact is visually apparent, gradual, and persistent. In the first six months following digitalization, households in the treated counties collectively buy, on average, 17.69 more flood insurance policies (corresponding to a 16.47% increase over the mean of 107.42 at event time zero). This gap continues widening to 32.84 (or 30.57%) at the end of the first year post digitalization. Thereafter, the number of policies in-force remains stable and never reverts. Recall that the policies are renewed annually. Thus, comparing event months 1-6 with event months 13-18, we could posit that the additional purchases made immediately after digitalization are renewed in the next year. To illustrate that the impact is not transitory, Appendix Figure A.4 extends the same plot to a horizon of 8 years and shows that the pattern persists in the long run.

4.2 First-time Buyers and Renewals

In this section, I investigate the surge in insurance purchases in more detail. I examine the number of policies purchased (a flow measure), complementing the previous analysis of the number of policies in-force (a stock measure). The main benefit of the flow measure is a clear decomposition of all purchases into two parts—first-time buyers and existing buyers (i.e., renewal decisions).

Figure 6.a depicts the dynamic effects of digitalization on the number of policies purchased. Similar to Figure 5, there is no evidence of a differential pre-trend in the treated counties. There is a sharp spike immediately following the publication of the digitalized maps—the transaction volume increases by 80 percent in event month 2. Interestingly, the dynamic effects display a recursive pattern. For instance, the second spike of insurance purchases occurs in event month 14, suggesting that the additional purchases are renewed one year later.

In the next step, I decompose the number of insurance purchases (time-specific) into policies bought by first-time buyers (Figure 6.b) and policies renewed by existing buyers (Figure 6.c). Note that these classifications of buyers are time-specific. Figure 6.b shows that the number of first-time buyers triples two months following digitalization. Noticeably, there is only one spike in Figure 6.b, as the first-time buyers in event month 2 become existing buyers by definition at the time of renewal in event month 14. This is consistent with the spike observed in Figure 6.c in that same month.

Crucially, Figure 6.c also suggests that digitalization has little impact on the extensive margin of existing policyholders, as we do not see households canceling or failing to renew their policies around event time zero. This finding has important implications for differentiating among alternative explanations (see details in Section 4.6).

4.3 Risk Map Digitalization and House Prices

The salience theory suggests that consumers place greater weight on product attributes that are salient (Bordalo et al., 2012, 2013; Kőszegi and Szeidl, 2013). In this section, I study a product that is closely related to flood risk—houses.

Conceptually, when flood risk is not salient, people pay limited attention to a property's flood risk exposure (a product attribute). As a result, this low-salience risk is not fully priced into real estate values. My hypothesis implies that when the digitalization of flood hazard maps increases its salience, we should expect it to be more heavily priced—that is, to drag down house prices.

Empirically, I examine the effect of digitalization on house prices in the same framework as regression (1). I obtain the monthly county-level home value index from Zillow (with January 2010 normalized to 100). Zillow provides separate indices for different types of homes, such as all homes, single-family homes, top-tier homes, bottom-tier homes, and homes with 1, 2, 3, 4, or 5+ bedrooms. For robustness, I also use the yearly county-level house price index from the Federal Housing Finance Agency (FHFA), in which case, I collapse my NFIP data on an annual basis as well.

Table 3 reports the coefficient estimates using the Zillow data. The baseline result, presented in column 1 of Panel A, suggests that the average treatment effect on house price index (all homes) is −1.86, corresponding to a 1.80% decrease relative to the mean value of 103.50 at event time zero. Column 2 includes covariates, such as the county-level median household income and unemployment rate. While they show significant explanatory power (unreported for brevity), the coefficient estimate of $digitalization_{it}$ is fairly unaffected. In column 3, I examine single-family homes only. The result is qualitatively and quantitatively similar.

Columns 4 and 5 demonstrate the heterogeneous effects of digitalization across the distribution of house prices. Zillow measures its house price index for the top-tier (bottom-tier) homes by using properties with values within the 65th to 95th (5th to 35th) percentile range for a given county. The results suggest that the impact of the introduction of digital maps is stronger for the cheaper properties, as the price drop for top-tier houses is

0.4 percentage point smaller than that for bottom-tier houses.

Panel B of Table 3 examines the house price indices of homes with various numbers of bedrooms. In all cases, there is a significant decrease after digitalization. Interestingly, the price drop is monotonically smaller in magnitude for houses with an increasing number of bedrooms. This result is consistent with the comparison in Panel A between the top-tier and bottom-tier homes.

Figure 7 shows the dynamic effects (using the baseline index of all homes). Firstly and crucially, there is little evidence of a differential pre-trend in the treated counties. Following digitalization, the house price index is apparently trending downwards. This result suggests that flood risk, once it attracts more public attention and has higher salience, is more extensively incorporated into asset prices. Finally, it is interesting to notice that unlike the immediate surge in flood insurance purchases (see Figure 5), the downward trend in house prices continues for a few years and then stabilizes around −2%. A plausible explanation is that housing transactions are much less liquid than insurance purchases.

For robustness, Appendix Figure A.5 presents the results of using the alternative annual house price index obtained from the FHFA. The downward post-digitalization trend has a similar pattern, and the estimate is larger in magnitude.

4.4 Salience and Past Flood Occurrence

As a flood would certainly draw the attention of those flooded and increase the salience of flood risk to them, my hypothesis predicts a smaller impact of digitalization if the county has recently experienced a flood. Moreover, the correlation should be somewhat monotonic—the longer the county has not experienced a flood, the stronger the digitalization effect should be.

I repeat my main analysis in subsamples of counties that have not had any flood in the n years prior to the publication of the digitalized maps. Panel A of Table 4 reports the results up to $n = 6$ (82% of the counties had at least one flood in the previous six years). I find an almost monotonic pattern going from columns 1 through 6. For instance, in counties without a flood in the previous year, digitalization causes the insurance demand to increase by about 20%, but in counties without a flood for at least six years, the impact of digitalization is twice as large.

As the frequency of flooding correlates with the inherent risk, it is worth showing that I do not merely replicate the variations in the underlying risk. To measure the inherent flood risk, I leverage the Special Flood Hazard Area (SFHA) defined by the NFIP, which is expected to have a one-percent or higher probability of being inundated in any given year. In Appendix Table A.3, I condition on counties with low flood risk (below-median proportions of SFHA), and I find the same monotonicity seen in Panel A of Table 4.

Panel B of Table 4 presents an alternative perspective to consider the correlation between salience, flood occurrence, and the effect of digitalization. In Panel B, I examine counties that had at least one flood in the past n years before digitalization. Given a value of n (i.e., fixing the column), my hypothesis predicts a larger effect in Panel A than in Panel B. The results, across all the columns, are consistent with this prediction.

4.5 Heterogeneity across Counties

In this section, I demonstrate the heterogeneity of my finding by examining several factors that pertain to households' attention to flood risk management. My hypothesis predicts that the post-digitalization increase in insurance purchases should be more substantial in counties where flood risk is more likely to be neglected.

4.5.1 Heterogeneity by Income and Education

I first consider income and education as a proxy for households' awareness of the importance of mitigating flood risk. To assess if people react differently to the digitalization reform, I interact the post-treatment indicator $Digitalization_{it}$ with $Income_i$ and $Education_i$ in regression (1). $Income_i$ measures county i's median household income (in \$1,000), and $Education_i$ measures the share (in percent) of population with college degrees in county i. The data is obtained from the US Census Bureau.

Columns 1 and 2 of Table 5 report the coefficient estimates of the interactions. The negative coefficients suggest that counties with more income and education are less responsive to the publication of the digitalized maps. Quantitatively, \$1,000 more in income or 1 percentage point more of college attainment is associated with 0.65 percentage point less growth in insurance purchases following digitalization.

My strategy controls for existing differences in levels of insurance purchases (e.g., the wealthier are more likley to acquire insurance in the first place), and I estimate the heterogeneous effect of the digitalization reform. My result suggests that inattention to flood risk is potentially more concerning in areas with less education and income.

4.5.2 Heterogeneity by Beliefs about Global Warming

Households' beliefs about global warming is another useful proxy for their ex-ante attention to flood risk. I obtain data from the Yale Climate Opinion survey (Howe et al., 2015), which has been used in several recent studies of climate risk (Bernstein et al., 2019; Baldauf et al., 2020). My main measure, $ClimateOpinion_i$, is the percentage of people, in county i, who answered "Yes" to the question of whether they think global warming will harm them personally.

Column 3 of Table 5 reports the coefficient estimate of $ClimateOpinion_i$ interacted with $Digitalization_{it}$. The result suggests that a 1-percentage-point increase in the proportion of people who worry about the consequence of global warming leads to a 1.56-percentage-point decrease in the post-digitalization growth in flood insurance purchases. Column 4 shows that this effect is not subsumed by the heterogeneity in income.

4.5.3 Heterogeneity by Inherent Risk Level

Because people living in areas with higher inherent flood risk are likely to pay more attention to it, the digitalization-induced salience shock should be less pronounced in flood-prone areas. Along this line, I consider several analyses.

First, I define a binary variable $\mathbb{1}(Coastal)_i$, equal to 1 if county i is located in a coastal state, as a proxy for high flood risk. Second, for every county, I calculate the proportion of policies originated in the Special Flood Hazard Area (SFHA) over the total number of policies and define $\mathbb{1}(HighRisk)_i$ as a binary variable indicating if county i is above the median. As before, I examine their interactions with $Digitalization_{it}$ to capture any heterogeneous treatment effect. The negative coefficients, reported in columns 5 and 6 of Table 5, suggest that the digitalization effect is stronger in areas with lower flood risk. Appendix Figure A.6 illustrates the point graphically and verifies that the parallel pre-trend assumption is satisfied in all subsamples.

Additionally, I exploit the heterogeneous effect within counties. My transaction data allows me to see if a policy was originated by a household living in the SFHA or not. I compare the post-digitalization growth in insurance purchases in a treated county's SFHA regions and non-SFHA regions. Columns 1 and 2 of Appendix Table A.4 suggest that the growth is less than 10 percent in the SFHA regions, but more than 50 percent in the non-SFHA regions. Column 3 effectively runs a triple-differences regression, showing that the difference is statistically significant. Consistently, column 4 shows that the proportion of SFHA policies decreases by 4.7 percentage points in the treated counties. These results suggest that the reform induces disproportionately more households in moderate- or low-risk areas to acquire flood insurance.

4.5.4 U-shaped Effect and Inherent Risk Level

Following up on the previous section, I explore the variation in flood risk in a more granular scope. The digitalization effect is likely to be non-linear in flood risk: on the one hand, in high-risk areas, the salience is already high; on the other hand, in low-risk areas, flood insurance is less relevant. Therefore, we expect to see the strongest response to the digitalization reform in moderate-risk areas—that is, an inverse U-shape.

In column 7 of Table 5, I consider a continuous measure $RiskLevel_i$, which is the proportion of SFHA policies in county i (in percent), interacting its quadratic form with the post-treatment dummy $Digitalization_{it}$. The result, in column 7, is consistent with the prediction: the estimated coefficients imply that the largest digitalization effect occurs when $RiskLevel_i$ equals 30 percent.

To illustrate this point graphically, I divide all counties into quintiles based on $RiskLevel_i$ and estimate the digitalization effect in each quintile. Figure 8.a plots the coefficient estimate of the effect of digitalization on the number of flood insurance policies in-force. Again, we observe an inverse U-shape. The moderate-risk counties (quintiles 2 and 3) experience the most substantial increase in insurance purchases following digitalization, whereas there is little response in the bottom and top quintiles.

To complement the above evidence supporting the salience hypothesis, I document the same pattern of heterogeneous effects in the post-digitalization housing markets. Figure 8.b plots the coefficient estimate of the effect of digitalization on house prices in quintiles of counties based on the flood risk measure $RiskLevel_i$. Consistently, we see a U-shape: the moderate-risk counties display the largest decrease in house prices, whereas there is little response in the bottom and top quintiles.

4.6 Alternative Explanations

In this section, I evaluate a number of alternative explanations for my main findings and present empirical evidence to characterize their relevance.

4.6.1 Non-causal Interpretations

I first address concerns pertaining to the validity of the assumption of quasi-random treatment. The observed positive relationship between digitalization and flood insurance demand might not be causal if both are driven by a third factor. One alternative explanation is that the digitalized maps are introduced across geographies that experience increasing underlying flood risk, which simultaneously drives up insurance purchases. In other words, the digitalization process might endogenously capture an existing trend in underlying flood risk and insurance demand. However, the parallel pre-trend, shown in Figure 5, refutes this explanation. It suggests that the staggered digitalization process does not capture any existing trends in demand.

However, even in the absence of a pre-trend, the reform could still be endogenous. One specific alternative mechanism is that the publication of digital maps follows a local flooding event, and households buy more flood insurance after recently experiencing a flood, due to either rational learning or behavioral bias. Many papers show that individuals overweight recent experiences (Greenwood and Nagel, 2009; Malmendier and Nagel, 2011; Dessaint and Matray, 2017), including floods (Gallagher, 2014). This explanation differs from the previous one if flooding is unpredictable, in which case the pre-trend test might not help. (For instance, two identical counties could have identical dynamics of flood insurance policies in-force, until a random one of them is hit by a flood.)

Therefore, I provide additional evidence to cast doubt on this alternative explanation. First, Appendix Table A.6 shows that past floods cannot predict the introduction of digital maps. Second, Appendix Figure A.7 shows that there is no difference in flooding frequency around the time of digitalization, although if this alternative hypothesis were true, we would expect to see bunching before it. Third, anecdotally, local officials explicitly declared that the publication of new digital maps is unrelated to recent local flooding events. For example, in an announcement posted on the official website of the City of Alexandria, Virginia, it says, "This effort [developing new digital maps] is unrelated to recent flooding the city has experienced from flash flooding in July 2019 and, more recently, on July 23 [2020]."[17]

4.6.2 Rational Explanation: Transaction Cost

After addressing the concerns of non-causal interpretations, I continue to explore if there is a purely rational alternative explanation for why the digitalization reform leads to increasing demand for flood insurance.

The first purely rational alternative explanation I consider is that the digitalization reform reduces the transaction cost of acquiring flood risk information. In other words, the observed marginal households did not buy flood insurance before digitalization because they found it too costly to search for information about flood risk, not because of a psychological cost related to inattention.

According to households' pre- and post-digitalization status, we can categorize four groups: (1) the insured (who already know their risk exposures as stated in their policies); (2) the uninsured who saw the paper maps before but decided not to buy flood insurance; (3) the uninsured who would not have seen the maps without digitalization due to a high transaction cost; and (4) the uninsured who ignore flood risk and any maps whatsoever.

In this alternative hypothesis, the digitalization does not affect group 1, 2, and 4, as it does not bring them any new information. Amongst group 3, assuming unbiased prior beliefs, some of these uninsured households who had previously underestimated their flood risk exposure would start buying flood insurance, thus accounting for my main finding; while the other households in group 3 would stay uninsured.

While my data does not allow me to completely rule out this alternative explanation, I present some evidence that it is unlikely to be the primary channel. First, the transaction cost of getting paper maps, although difficult to quantify, was likely minimal. Households could always obtain the scanned copies of paper maps from FEMA's website, and it was likely not a hassle to visit their local NFIP offices or call its toll-free hotlines.

Second, as discussed above, under this alternative hypothesis, existing policyholders (i.e., group 1) should not behave differentially before and after digitalization. However, I find that digitalization also affects the intensive margin of the already insured households: they increase coverage per policy by 2.6 percent following digitalization.

[17] See https://www.alexandriava.gov/floodmap for full information.

Third, this alternative explanation does not imply that the digitalization effect should be stronger in moderate-risk areas than in high-risk areas, yet that is what I found, as discussed in Sections 4.5.2 and 4.5.3.

Fourth, I consider an additional quasi-experiment featuring a reduction in transaction cost. Appendix Section B details the empirical setting, but in a nutshell, in July 2014, the government upgraded its online portal to provide improved search functionalities and greater convenience to users. I test whether this upgrade stimulates insurance demand. Specifically, I run a standard difference-in-differences regression, where the treated are counties with digital maps published before July 2014, and the control are counties that had not implemented digitalization. The result, reported in Appendix Table A.5, suggests little impact on insurance purchases.

4.6.3 Rational Explanation: Change in Informational Content

The second rational alternative explanation I consider is that the informational content of the paper and digital maps is fundamentally different and that the observed increase in insurance purchases merely reflects Bayesian updating. This is a valid concern, as anecdotally some maps may have been updated while being digitalized.

However, I present evidence that the scope of this alternative channel is small. First, the Web Appendix presents many examples comparing the digital and paper maps. Visually, the flood risk information in the two formats appears to be almost identical. I also find consistent anecdotal evidence; for example, a press release for the new digital maps for Sussex County, New Jersey, quotes Mary Colvin, Acting Mitigation Director for FEMA, Region II, as saying: "The new, preliminary map does not present any major changes in the flood plain."[18]

Second, the government suggests that the update component can go either way. Specifically, for the purpose of my argument, some areas should be downgraded to median or low-risk, as "overstated hazards can result in potentially unnecessary construction costs and incorrect insurance rating decisions." If this alternative explanation is prevailing, we should observe some households cancel or fail to renew their policies (even if the update is possibly systematically upwards—that is, if disproportionately more areas are designated with a higher risk). However, my result, presented in Section 4.2, suggests

[18] See the county's website: https://www.sussex.nj.us/cn/news/index.cfm?TID=7&NID=17552.

that digitalization does not have any significant impact on renewals or cancellations.

Third, I show, in Section 4.5.3, that the post-digitalization effect is stronger in areas with lower inherent flood risk. To reconcile this result, the alternative mechanism needs to assume that the systematic increases in flood risk (reflected in the new maps) are disproportionately more significant in lower-risk areas. To the best of my knowledge, this assumption is not anecdotally supported.

Fourth, using a complementary sample (only yearly) of the program from 1980 to 2000, I find that map updates (from paper to paper) do not stimulate insurance purchases.[19] It suggests that FEMA's map modifications are likely to be small in scope.

4.6.4 Changes in Insurance Price

Is it possible that the introduction of digital maps comes along with reductions in insurance price? This alternative hypothesis is not possible, as the federal government sets the NFIP rates and prohibits price discrimination across localities. Therefore, county officials and agents have no authorization to amend insurance prices while introducing the new digital maps.[20]

Nevertheless, to empirically evaluate this alternative explanation, I check whether there are any price differences before and after digitalization. I consider three county-level measures of insurance price: the average premium per policy, the average premium per policy per $1,000 coverage, and the average premium per policy scaled by the proportion of high-risk policies (to account for changes in risk composition). I also examine the county-level median household income which pertain to the real cost of purchasing flood insurance. Appendix Figure A.8 shows that there is no evidence to support this alternative explanation.

5 Empirical Findings: Second Strategy

In this section, I present the findings of my second quasi-experiment, developed in Section 3.2. It complements the results of my first strategy and supports the hypothesis that

[19] The transaction-level information from the NFIP is available only for after 2009. For the earlier periods, the program provides the year-end statistics of the number of policies in-force for all counties.

[20] There were two nationwide reforms regarding insurance rates: the Biggert-Waters Flood Insurance Reform Act of 2012 and the Homeowner Flood Insurance Affordability Act of 2014. These affect all counties simultaneously and are controlled for by my difference-in-differences strategy.

households' flood insurance decisions are sensitive to the salience of flood risk.

5.1 Social Connectedness and Geographically Distant Floods

I exploit variations in social connectedness on Facebook and leverage non-local floods. The main idea is that a geographically distant flood should not change or generate any information about the actual local flood risk. However, non-local flood news potentially conveys differential degrees of salience to the local households, depending on the strength of social connectedness.

Using the event study framework of regression (4), I compare the changes in flood insurance policies in-force across counties in the same state with high versus low levels of social connectedness to a geographically distant flooded area. The connected (treated) counties are defined by having a connectedness measure above the state median.

Table 6 reports the coefficient estimates of regression (4). To reject the null hypothesis, we expect a positive difference-in-differences estimate. In column 1 of Panel A, the estimate of 1.11 is a 0.94-percent increase over the average number of policies in-force (117.66) at event time zero. It suggests that, when individuals see geographically distant Facebook friends sharing flood experiences, increasing the salience of flood risk, more households decide to acquire flood insurance.

This result is robust to a variety of alternative specifications and sample restrictions. For example, column 2 uses an equal-weighting scheme in equation (3) to measure a county's social connectedness with the flooding counties. Columns 3 and 4 construct the analysis sample using counties that are at least 500 miles or 1,000 miles away from a given flood. The effect remains positive and statistically significant.

In Panel B of Table 6, I further check the robustness of my results. Because the Facebook data is a snapshot as of April 2016, one concern is that the county-by-county social connectedness might be time-varying. In particular, the measure might not be a good proxy for the status back in the early 2010s. Panel B of Table 6 uses a sample period between 2014 and 2017. My findings are not sensitive to the choices of sample periods.

Figure 9 presents the dynamic effects of the flood experiences of geographically distant friends. Most crucially, there is no evidence of a differential pre-trend in the connected counties before the flood, which supports the identifying assumption that the less-connected counties are an appropriate counterfactual for the more-connected counties in the same state.

5.2 Salience and Heterogeneity

5.2.1 Monotonicity in Social Connectedness

The salience hypothesis implies that the effect should be monotonic in the strength of social connectedness. The most-connected counties should show the largest increases in insurance demand and the least-connected counties should show the least change.

Recall that my baseline analysis (in Table 6) defines the treatment or control group as respectively above or below the state-median value of the social connectedness measure as per equation (3). In Panel A of Table 7, I use a sharper approach and compare the top versus bottom quartiles. The estimates are larger in magnitude across all specifications. For example, the estimate of 2.20 in column 1 is a 1.88-percent increase over the mean of 117.09 at event time zero. This suggests that the salience of flood news, transmitted through social networks, is monotonic in the strength of social connectedness.

5.2.2 Significant Floods

My hypothesis also implies that the most damaging floods should cause the most pronounced salience effect across social networks. Panel B of Table 7 tests this prediction. I restrict my event study to 18 floods that were characterized as significant by FEMA.[21] Across all specifications, the estimate is more than twice as large in magnitude as the baseline in Table 6. For example, the estimate of 3.08 in column 1 is a 2.64-percent increase over the mean of 116.73 at event time zero; column 4 generates the largest estimate (5.10%) in this empirical design. This result suggests that a natural disaster's salience, transmitting across social networks, is order-preserving.

5.3 Limitation and Alternative Methodology

I use an alternative methodology to address a concern of my event-study design to link social connectedness with flood insurance purchases. The advantage of my second strat-

[21] See https://www.fema.gov/significant-flood-events for FEMA's list of significant flood events. A significant event is defined as a flooding event with 1,500 or more paid losses.

egy is that each flooding event characterizes a standard difference-in-differences analysis, which allows for a straightforward verification of parallel pre-trends. However, the disadvantage is that a county could be involved (as either treated or control) in more than one events, which entails duplicating observations if the event windows overlap. In other words, instead of one observation per county per time, the event study approach has one observation per county per event per time.

I first address this problem of non-independent observations by clustering the standard errors at the county level (as in Tables 6 and 7). The other approach commonly adopted by empirical researchers is to only use large events, in the hope that they are sufficiently far apart. The analysis presented in Panel B of Table 7, with only the largest 18 floods included, is undertaken in this spirit.

In the following, I consider an alternative empirical approach used by Bailey et al. (2018a). Applying their terminology to my setting, I construct an index, $FriendFlood^{N}_{i,t_1,t_2}$, to measure the average flood experience of county i's social network N between t_1 and t_2. The largest social network N is the universe of all other counties; a restricted network can include only geographically distant ones. Let $\theta^{N}_{i,j}$ be the share of county i's friends in network N who live in county j, and let $Flood_{j,t_1,t_2}$ be the number of floods in county j between t_1 and t_2. The key explanatory variable is constructed as:

$$FriendFlood^{N}_{i,t_1,t_2} = \sum_{j} \theta^{N}_{i,j} * Flood_{j,t_1,t_2} \tag{5}$$

Bailey et al. (2018a) instrument for the house price experiences of all friends with the house price experiences of geographically distant friends to identify the causal impact of friends on an individual's housing investment decisions. As my primary interest is on the distant floods, I focus on the reduced-form to capture the average effect of geographically distant friends. Specifically, I estimate the following regression, with my baseline specification taking t_1 to be 12 months before t_2:

$$log(Policies)_{i,t} = \beta * FriendFlood^{Distant}_{i,t-12,t} + FE_{state\times time} + \epsilon_t \tag{6}$$

I control for the state×time fixed effects, which allow me to isolate the effects of friends' flood experiences on the insurance decisions of counties located in the same state at the same time.

Panel A of Table 8 presents results from regressions (6). The estimate in column 1 suggests that every flood experienced by friends living in geographically distant counties (at least 750 miles away) increases the local insurance demand by 1.3%. Columns 2 through 4 show that my result is robust to a variety of specifications with different measurement windows of floods. Columns 5 through 7 show that my result is also insensitive to a variety of definitions of "geographically distant".

Panel B of Table 8 repeats the analysis by focusing on the experiences of significant floods only (as defined in Section 5.5.2). Across all columns, the estimates in Panel B are more than twice the baseline in Panel A. For example, the estimate in column 1 means that when geographically distant friends experience a significant flood, the local county's demand for flood insurance increases by 4.2%. Consistent with my event study methodology, these findings suggest that a distant flood's salience effect, transmitting across social networks, is order-preserving.

6 Conclusion

This paper examines how households make insurance decisions against flood risk. I use two empirical strategies to identify a causal effect of the salience of flood risk on households' willingness to acquire flood insurance. My results suggest that households pay limited attention to flood risk, due to its low salience. But because the expected cost of neglecting flood risk is large, US policymakers are seeking ways to stimulate insurance take-up. My findings suggest that one effective way is to increase the salience of flood risk by running campaigns to enhance public awareness, presenting risk information in more salient formats to households, and covering non-local flood news on local media. This insight could be widely generalized to other types of tail risk (especially natural disaster risk) and to other countries.

References

S. Andersen, J. Y. Campbell, K. M. Nielsen, and T. Ramadorai. Sources of inaction in household finance: Evidence from the danish mortgage market. *American Economic Review*, 2020.

A. Anderson and D. T. Robinson. Knowledge, fear and beliefs: Understanding household demand for green investments. *Swedish House of Finance Research Paper*, 2019.

M. Bailey, R. Cao, T. Kuchler, and J. Stroebel. The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy*, 126(6):2224–2276, 2018a.

M. Bailey, R. Cao, T. Kuchler, J. Stroebel, and A. Wong. Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3):259–80, 2018b.

M. Baldauf, L. Garlappi, and C. Yannelis. Does climate change affect real estate prices? only if you believe in it. *The Review of Financial Studies*, 33(3):1256–1295, 2020.

A. Banerjee, A. Finkelstein, R. Hanna, B. A. Olken, A. Ornaghi, and S. Sumarto. The challenges of universal health insurance in developing countries: Evidence from a large-scale randomized experiment in indonesia. Technical report, National Bureau of Economic Research, 2019.

B. M. Barber and T. Odean. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies*, 21(2):785–818, 2008.

G. Bernile, V. Bhagwat, and P. R. Rau. What doesn't kill you will only make you more risk-loving: Early-life disasters and ceo behavior. *The Journal of Finance*, 72(1): 167–206, 2017.

A. Bernstein, M. T. Gustafson, and R. Lewis. Disaster on the horizon: The price effect of sea level rise. *Journal of Financial Economics*, 134(2):253–272, 2019.

P. Bordalo, N. Gennaioli, and A. Shleifer. Salience theory of choice under risk. *The Quarterly journal of economics*, 127(3):1243–1285, 2012.

P. Bordalo, N. Gennaioli, and A. Shleifer. Salience and consumer choice. *Journal of Political Economy*, 121(5):803–843, 2013.

D. Bose, H. Cordes, S. Nolte, J. C. Schneider, and C. F. Camerer. Decision weights for experimental asset prices based on visual salience. *Working Paper*, 2020.

J. Brown, T. Hossain, and J. Morgan. Shrouded attributes and information suppression: Evidence from the field. *The Quarterly Journal of Economics*, 125(2):859–876, 2010.

M. R. Busse, D. G. Pope, J. C. Pope, and J. Silva-Risso. The psychological effect of weather on car purchases. *The Quarterly Journal of Economics*, 130(1):371–414, 2015.

T. Y. Chang, W. Huang, and Y. Wang. Something in the air: Pollution and the demand for health insurance. *The Review of Economic Studies*, 85(3):1609–1634, 2018.

R. Chetty, A. Looney, and K. Kroft. Salience and taxation: Theory and evidence. *American economic review*, 99(4):1145–77, 2009.

D. Choi, Z. Gao, and W. Jiang. Attention to global warming. *The Review of Financial Studies*, 33(3):1112–1145, 2020.

J. J. Choi, D. Laibson, B. C. Madrian, and A. Metrick. Reinforcement learning and savings behavior. *The Journal of finance*, 64(6):2515–2534, 2009.

S. Cole, X. Giné, J. Tobacman, P. Topalova, R. Townsend, and J. Vickery. Barriers to household risk management: Evidence from india. *American Economic Journal: Applied Economics*, 5(1):104–35, 2013.

M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara. Predicting human eye fixations via an lstm-based saliency attentive model. *IEEE Transactions on Image Processing*, 27 (10):5142–5154, 2018.

O. Dessaint and A. Matray. Do managers overreact to salient risks? evidence from hurricane strikes. *Journal of Financial Economics*, 126(1):97–121, 2017.

A. Dittmar and R. Duchin. Looking in the rearview mirror: The effect of managers' professional experience on corporate financial policy. *The Review of Financial Studies*, 29(3):565–602, 2016.

A. Finkelstein. E-ztax: Tax salience and tax rates. *The Quarterly Journal of Economics*, 124(3):969–1010, 2009.

A. Finkelstein, N. Hendren, and E. F. Luttmer. The value of medicaid: Interpreting results from the oregon health insurance experiment. *Journal of Political Economy*, 127(6):2836–2874, 2019.

J. Gallagher. Learning about an infrequent event: evidence from flood insurance take-up in the united states. *American Economic Journal: Applied Economics*, pages 206–233, 2014.

S. Giglio, M. Maggiori, J. Stroebel, and A. Weber. Climate change and long-run discount rates: Evidence from real estate. Technical report, National Bureau of Economic Research, 2015.

R. Greenwood and S. Nagel. Inexperienced investors and bubbles. *Journal of Financial Economics*, 93(2):239–258, 2009.

J. S. Hastings and J. M. Shapiro. Fungibility and consumer choice: Evidence from commodity price shocks. *The quarterly journal of economics*, 128(4):1449–1498, 2013.

T. L. Hayes and D. A. Neal. Actuarial rate review: In support of the recommended october 1, 2011, rate and rule changes. *Washington, DC: Federal Emergency Management Agency*, 2011.

D. P. Horn and B. Webel. Private flood insurance and the national flood insurance program. *Congressional Research Service*, 2019.

P. D. Howe, M. Mildenberger, J. R. Marlon, and A. Leiserowitz. Geographic variation in opinions on climate change at state and local scales in the usa. *Nature Climate Change*, 5(6):596–603, 2015.

P. Issler, R. Stanton, C. Vergara-Alert, and N. Wallace. Mortgage markets with climate-change risk: Evidence from wildfires in california. *Available at SSRN 3511843*, 2019.

D. Karlan, R. Osei, I. Osei-Akoto, and C. Udry. Agricultural decisions after relaxing credit and risk constraints. *The Quarterly Journal of Economics*, 129(2):597–652, 2014.

J. M. Keenan, T. Hill, and A. Gumber. Climate gentrification: from theory to empiricism in miami-dade county, florida. *Environmental Research Letters*, 13(5):054001, 2018.

B. Kőszegi and A. Szeidl. A model of focusing in economic choice. *The Quarterly journal of economics*, 128(1):53–104, 2013.

T. Kuchler and B. Zafar. Personal experiences and expectations about aggregate outcomes. *The Journal of Finance*, 74(5):2491–2542, 2019.

N. Lacetera, D. G. Pope, and J. R. Sydnor. Heuristic thinking and limited attention in the car market. *American Economic Review*, 102(5):2206–36, 2012.

X. Li and C. Camerer. Using visual salience in empirical game theory. *Available at SSRN 3308886*, 2019.

Y. Li, E. J. Johnson, and L. Zaval. Local warming: Daily temperature change influences belief in global warming. *Psychological science*, 22(4):454–459, 2011.

U. Malmendier and Y. H. Lee. The bidder's curse. *American Economic Review*, 101(2): 749–87, 2011.

U. Malmendier and S. Nagel. Depression babies: do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics*, 126(1):373–416, 2011.

U. Malmendier and S. Nagel. Learning from inflation experiences. *The Quarterly Journal of Economics*, 131(1):53–87, 2016.

U. Malmendier, G. Tate, and J. Yan. Overconfidence and early-life experiences: the effect of managerial traits on corporate financial policies. *The Journal of finance*, 66 (5):1687–1733, 2011.

J. Murfin and M. Spiegel. Is the risk of sea level rise capitalized in residential real estate? *The Review of Financial Studies*, 33(3):1217–1255, 2020.

A. Ouazad and M. E. Kahn. Mortgage finance in the face of rising climate risk. Technical report, National Bureau of Economic Research, 2019.

A. Schoar and L. Zuo. Shaped by booms and busts: How the economy impacts ceo careers and management styles. *The Review of Financial Studies*, 30(5):1425–1456, 2017.

V. Stango and J. Zinman. Limited and varying consumer attention: Evidence from shocks to the salience of bank overdraft fees. *The Review of Financial Studies*, 27(4):990–1030, 2014.

S. E. Taylor and S. C. Thompson. Stalking the elusive" vividness" effect. *Psychological review*, 89(2):155, 1982.

R. B. Towal, M. Mormann, and C. Koch. Simultaneous modeling of visual saliency and value computation improves predictions of economic choice. *Proceedings of the National Academy of Sciences*, 110(40):E3858–E3867, 2013.

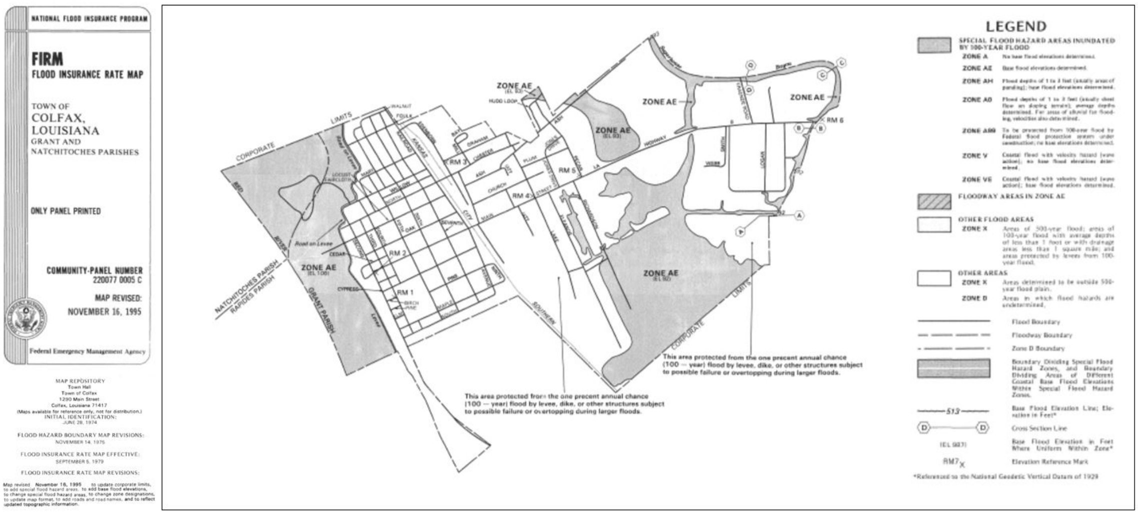
A. Tversky and D. Kahneman. Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2):207–232, 1973.

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

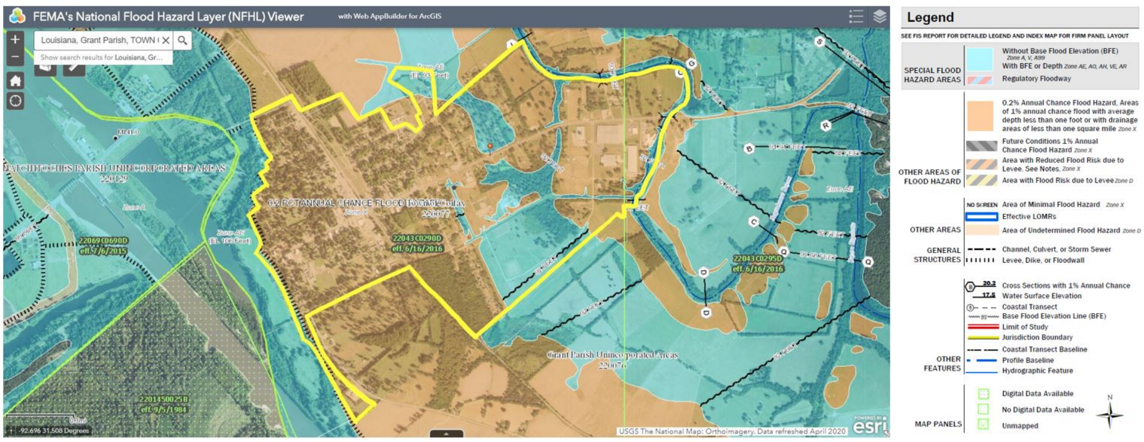
K. Wagner. Adaptation and adverse selection in markets for natural disaster insurance. *Available at SSRN 3467329*, 2019.

L. Zaval, E. A. Keenan, E. J. Johnson, and E. U. Weber. How warm days increase belief in global warming. *Nature Climate Change*, 4(2):143–147, 2014.

Figures and Tables



(a) Paper Map Published on November 16, 1995

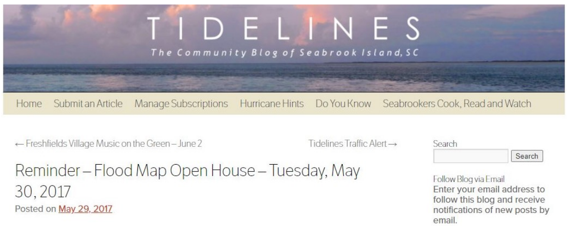


(b) Digital Map Published on June 16, 2016

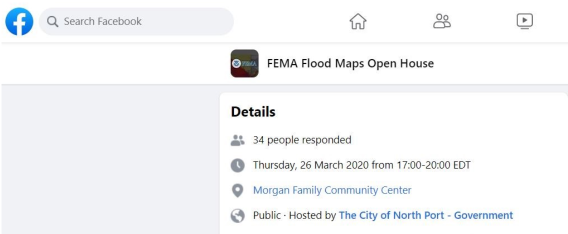
Figure 1. An Example of a National Flood Insurance Program Flood Hazard Map

This figure shows the flood hazard maps developed by the National Flood Insurance Program for the Town of Colfax, Grant Parish, Louisiana. Figure (a) is a scanned copy of the legacy paper map, which was published on November 16, 1995. For readability, only the most relevant information is presented here, and the full copy can be found at https://msc.fema.gov/portal. Figure (b) shows the corresponding digital map published on June 16, 2016. The two maps present identical information about the flood risk in the Town of Colfax (except that the jurisdiction boundary is slightly different).

(a) Open House Invitation on Local News

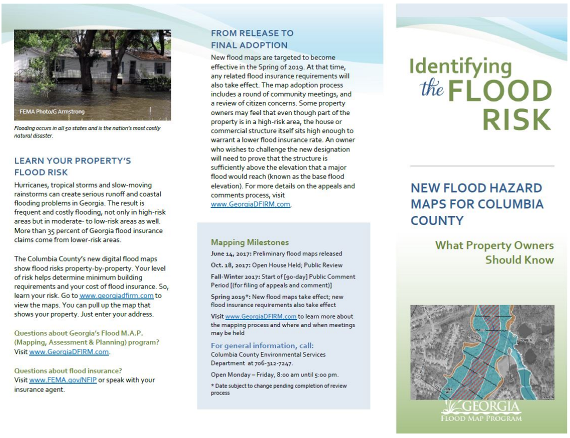


(b) Advertisement on Community Blog



(c) Open House Invitation via Facebook

(d) Announcement of New Maps Publication



(e) Brochure

(f) Local Newspaper

Figure 2. Local Advertisements of Flood Risk Open Houses and Map Publication

This figure presents examples of county governments advertising Flood Risk Open Houses and the publication of new flood maps.

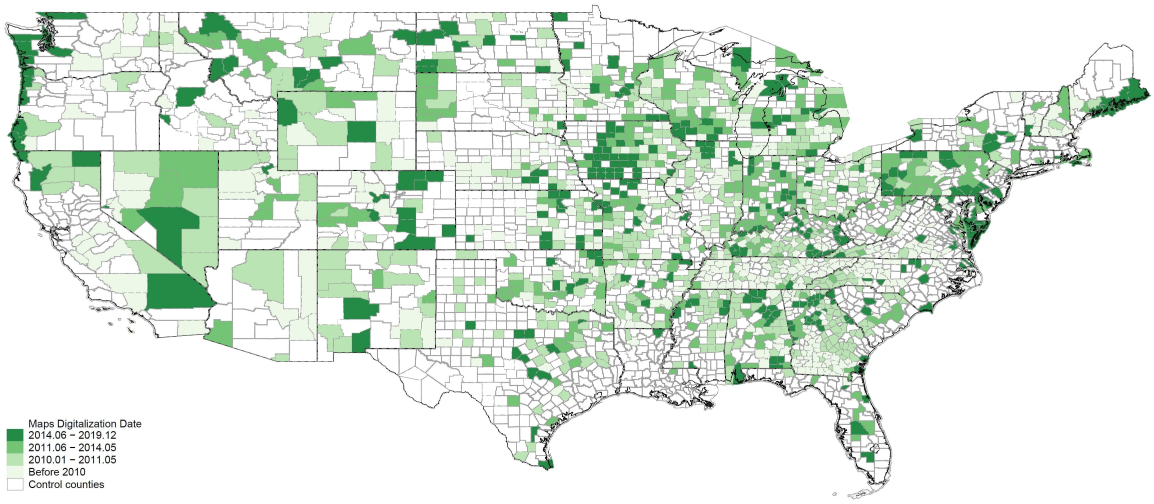


Figure 3. Empirical Strategy 1: Staggered Digitalization of Flood Risk Maps

The figure shows the rollout of the digitalization program by county and time. The darker shade represents the more recent publication date of the digitalized maps. The unshaded counties represent the untreated group.

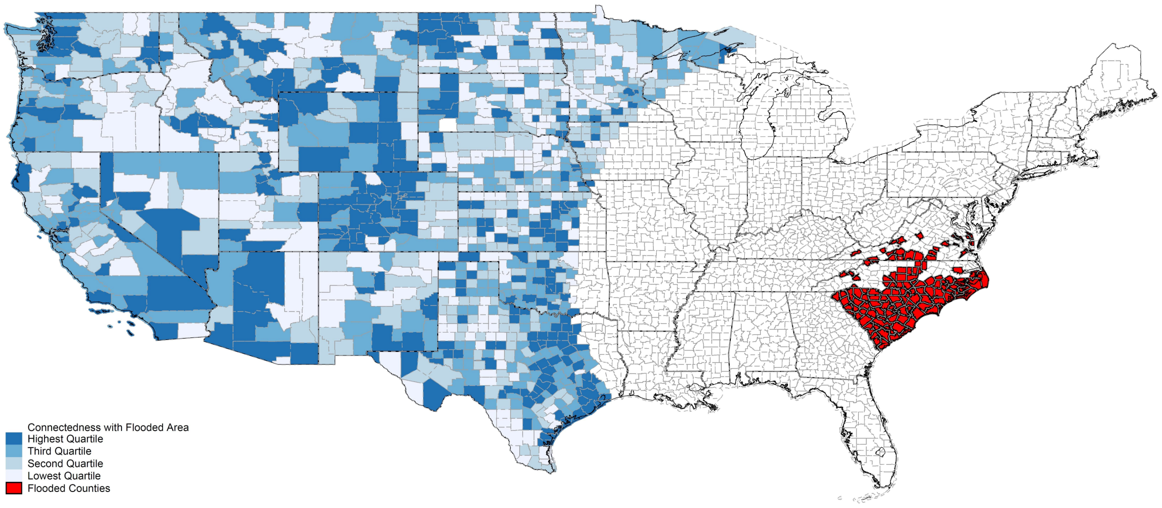


Figure 4. Empirical Strategy 2: Social Connectedness and Geographically Distant Floods

This figure shows one specific example to illustrate the empirical design of the second quasi-experiment. Hurricane Florence hit South Carolina, North Carolina, and Virginia in September 2018. The flooding counties are red-shaded on the map. The blue shades depict the heat map of social connectedness with the flooding area. Only counties located at least 750 miles away from the flooding area are considered.

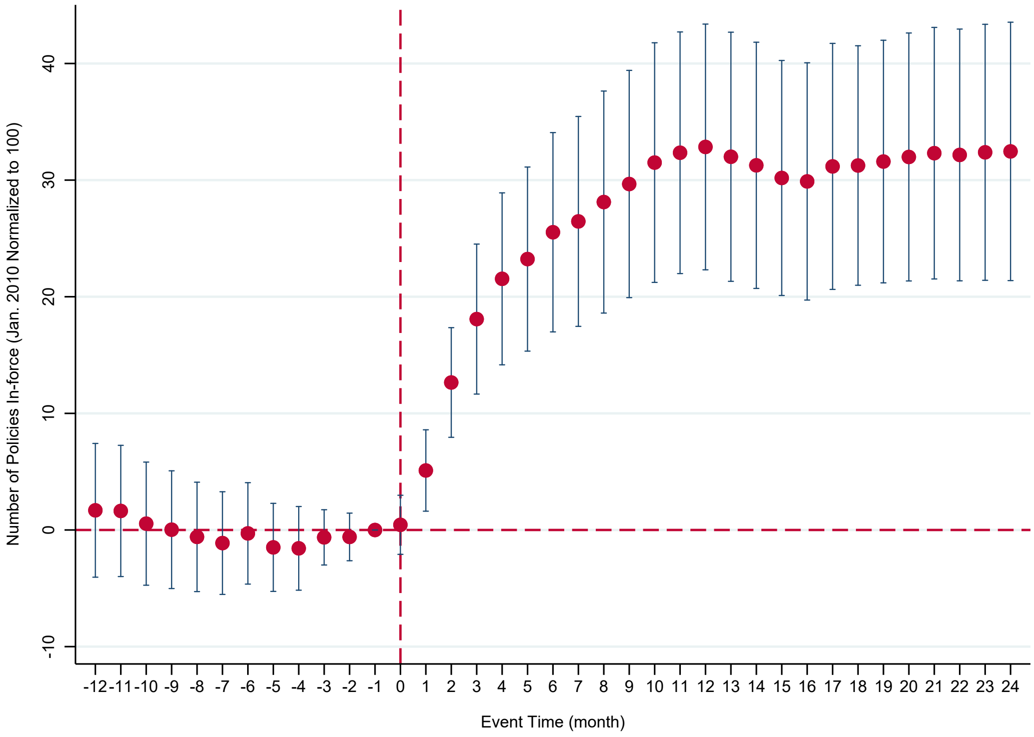
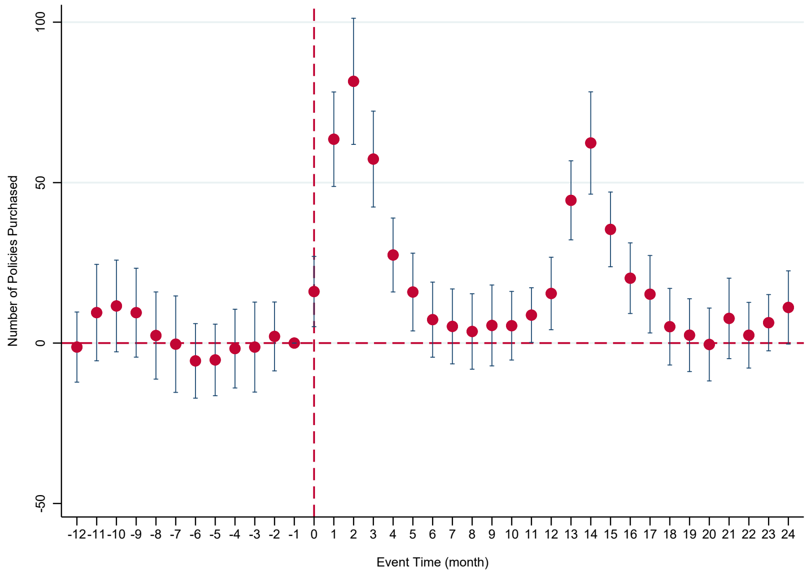
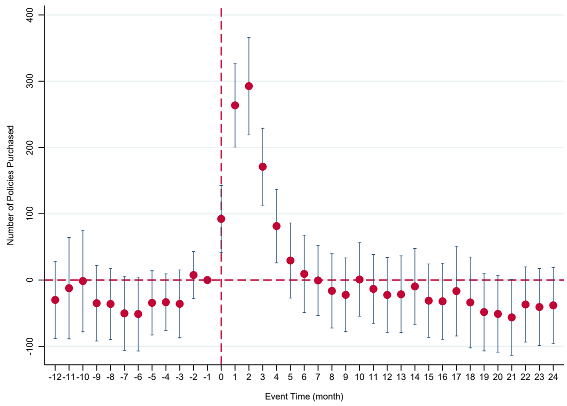


Figure 5. The Impact of Map Digitalization on the Outstanding Insurance Policies

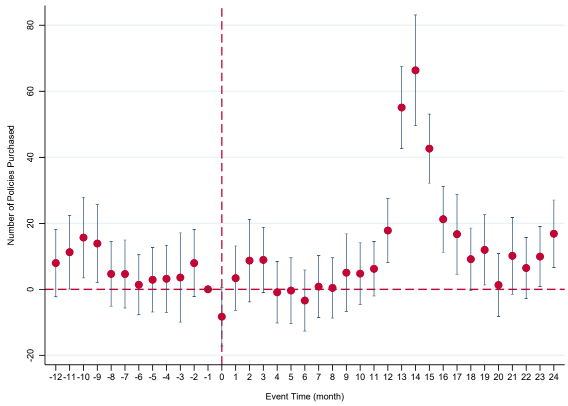
This figure shows the dynamic effects of risk maps digitalization on insurance purchases. It plots the coefficient estimates of $\{\beta_k\}$ in the regression: $Y_{it} = \alpha_i + \lambda_t + \sum_k \beta_k * \mathbb{1}(t = t_i^* + k) + \epsilon_{it}$. $\{\beta_k\}$ are measured relative to β_{-1} which is omitted. The dependent variable Y_{it} measures the number of flood insurance policies in-force (with January 2010 normalized to 100) in county i in month t. t_i^* is the publication time of the digitalized maps in county i. $\mathbb{1}(t = t_i^* + k)$ is set to zero for the untreated. α_i and λ_t are the county and year-month fixed effects. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.



(a) All Purchases



(b) First-time Buyers (time-specific)



(c) Existing Buyers (time-specific)

Figure 6. The Impact of Map Digitalization on the Insurance Purchases (Flow Measure)

This figure plots the coefficient estimates of $\{\beta_k\}$ in the regression: $Y_{it} = \alpha_i + \lambda_t + \sum_k \beta_k * \mathbb{1}(t = t_i^* + k) + \epsilon_{it}$. In figure (a), Y_{it} measures the number of flood insurance policies purchased in county i in month t. In figure (b), Y_{it} measures the number of flood insurance policies purchased by first-time buyers in county i in month t. In figure (c), Y_{it} measures the number of flood insurance policies renewed by existing buyers in county i in month t. In all cases, Y_{it} is normalized with the value of January 2010 being 100. All the other variables are defined as per Figure 3. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.

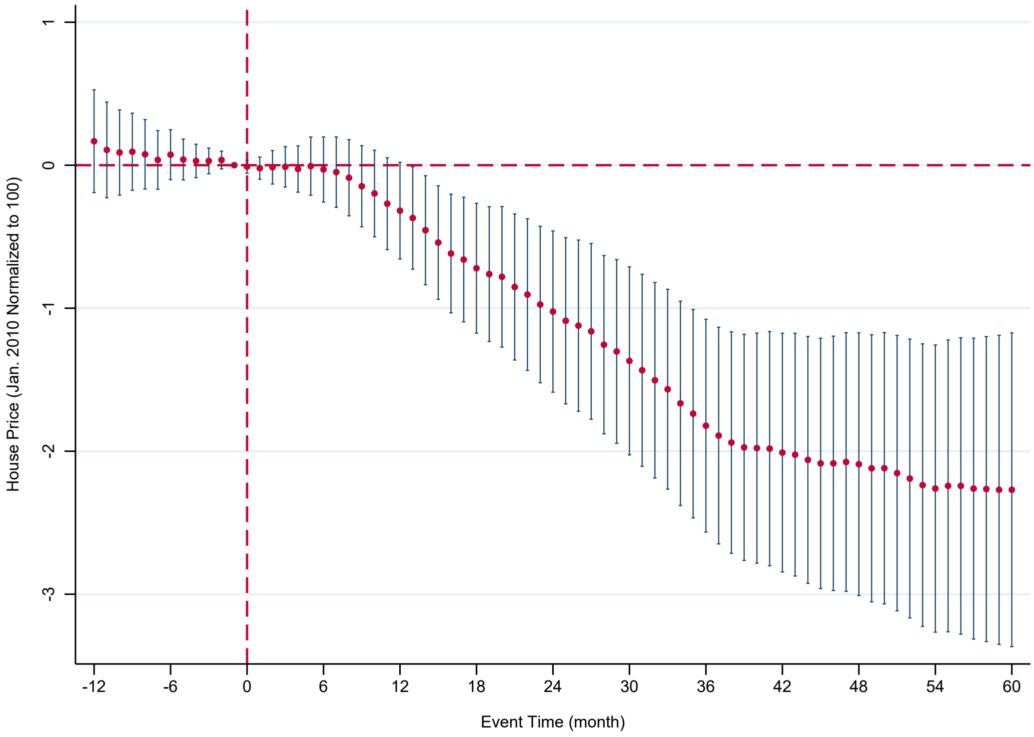
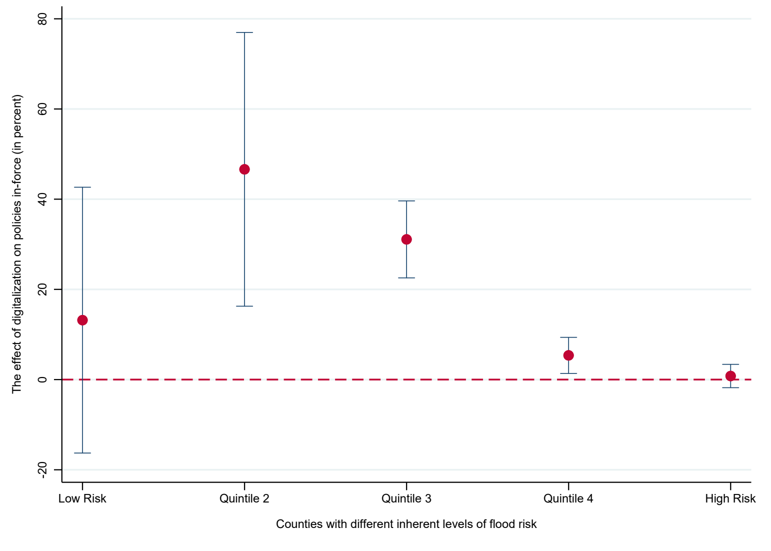
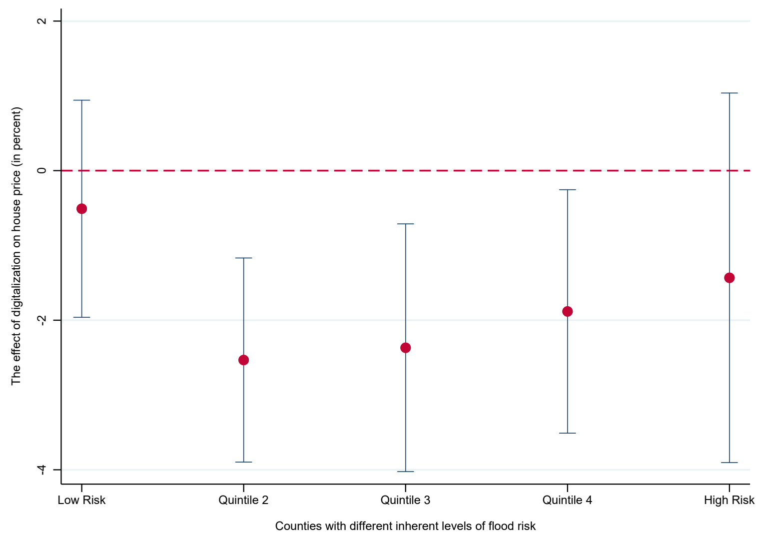


Figure 7. The Impact of Risk Maps Digitalization on Housing Price

This figure shows the dynamic effects of risk maps digitalization on house prices. It plots the coefficient estimates of $\{\beta_k\}$ in the regression: $HousePrice_{it} = \alpha_i + \lambda_t + \sum_k \beta_k * \mathbb{1}(t = t_i^* + k) + \epsilon_{it}$. $\{\beta_k\}$ are measured relative to β_{-1} which is omitted. The dependent variable $HousePrice_{it}$ is the house price index (with January 2010 normalized to 100) in county i in month t. t_i^* is the calendar month when county i publishes its digitalized maps. $\mathbb{1}(t = t_i^* + k)$ is set to zero for the untreated. α_i and λ_t are the county and year fixed effects. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.



(a) Heterogeneous effect on insurance purchases by risk-quintiles



(b) Heterogeneous effect on house prices by risk-quintiles

Figure 8. The Impact of Map Digitalization on the Insurance Purchases (Flow Measure)

This figure shows the coefficient estimate of $Digitalization_{it}$ in different subsamples from the regression: $Y_{it} = \alpha_i + \lambda_t + \beta * Digitalization_{it} + \epsilon_{it}$. All counties are divided into quintiles based on a measure of inherent flood risk $RiskLevel_i$, which is the proportion of Special Flood Hazard Areas (SFHA). In figure (a), the dependent variable Y_{it} measures the number of flood insurance policies in-force (with January 2010 normalized to 100) in county i in month t. In figure (b), Y_{it} measures the average house prices (with January 2010 normalized to 100) in county i in month t. The main explanatory variable $Digitalization_{it}$ is a binary variable indicating if county i has published the digitalized maps at time t. α_i and λ_t are the county and year-month fixed effects. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 90% confidence intervals.

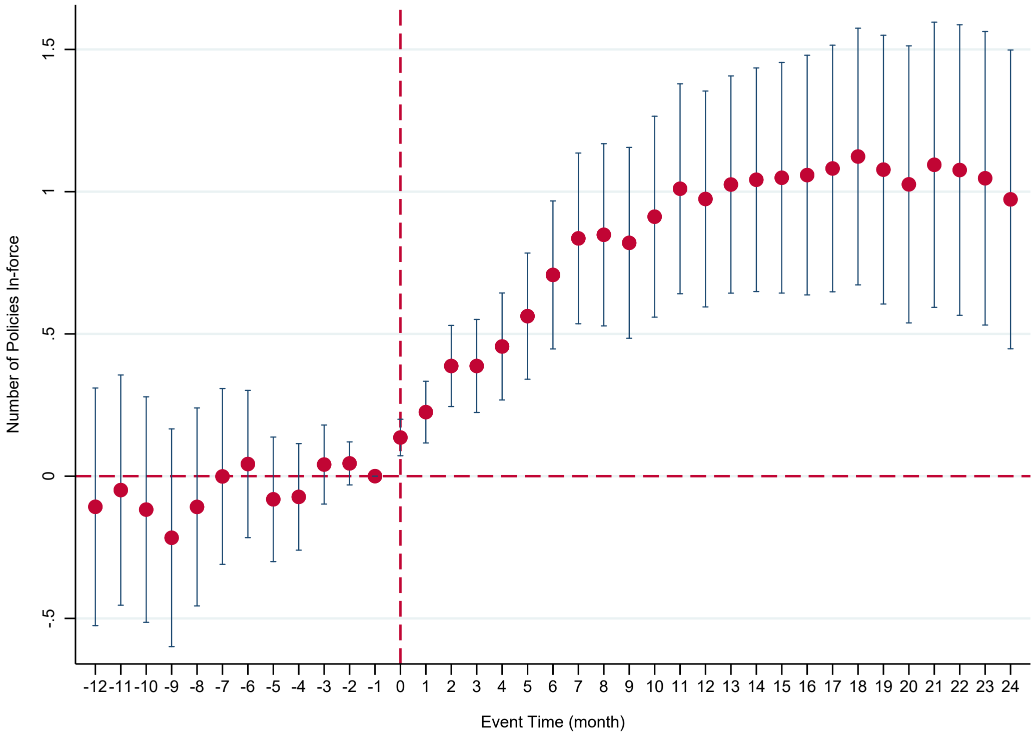


Figure 9. The Impact of Friends' Flood Experiences on Insurance Purchases

This figure shows the dynamic effects of geographically distant friends' flooding experiences on insurance purchases. It plots the coefficient estimates of $\{\beta_k\}$ from the event study design: $Y_{it} = \beta_0 + \sum_k \beta_1^k * Connected_i \times \mathbb{1}(t = t^* + k) + \beta_2 * Connected_i + \sum_k \beta_3^k * \mathbb{1}(t = t^* + k) + \epsilon_{it}$. For notational brevity, the event index f is omitted from the equation. $\{\beta_1^k\}$ are measured relative to $\beta_1^{k=-1}$ which is omitted. For a given flood event f and the associated flooding counties $\{j\}_f$, $Connected_i$ is a binary variable indicating if county i is socially connected with the flooding area, which is defined as having a value of the connectedness measure above the state-median. t^* is the occurrence month of the geographically distant flood. The analysis sample consists of only counties that are at least 750 miles away from the flooding area. The dependent variable Y_{it} measures the number of flood insurance policies in-force (with January 2010 normalized to 100) in county i in month t. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.

Table 1: Descriptive Statistics of the NFIP

This table presents descriptive statistics that characterize the National Flood Insurance Program (NFIP) data in my sample from January 2010 to August 2019. Panel A summarizes the time-series variations for the entire program. *Policies in-force* is the number of effective insurance policies in a given month. *Premium* is the total dollar amount of premiums collected from the policies in-force in a given month. *Coverage* is the total dollar amount of coverages for the policies in-force in a given month. *Flood-prone policies* is calculated as the number of policies in the Special Flood Hazard Area (SFHA) over the total number of policies. The NFIP creates risk maps and designates flood zones, and the SFHA is defined as the area that has a 1-percent or higher probability to be inundated in any given year. Panel B summarizes the cross-sectional variations of the data at the county level.

	Panel A: Nationwide Time-Series Varation				
	mean	s.d.	25^{th} pctl.	50^{th} pctl.	75^{th} pctl.
Policies in-force (m)	5.29	0.22	5.08	5.32	5.51
Premium (\$b)	3.32	0.13	3.26	3.30	3.41
Coverage (\$t)	1.26	0.03	1.24	1.26	1.28
Premium per policy (\$)	628	35.8	599	647	654
Coverage per policy (\$k)	238	12.1	229	239	249
Flood-prone policies (%)	52.2	3.8	49.6	52.9	55.7
	Panel B: County-level Cross-Sectional Variation				
	mean	s.d.	25^{th} pctl.	50^{th} pctl.	75^{th} pctl.
Policies in-force	1,766	12,563	31	120	437
Premium (\$k)	1,108	6,251	21	88	325
Coverage (\$m)	421	3,023	4.8	20	83
Premium per policy (\$)	754	358	554	697	876
Coverage per policy (\$k)	185	68	137	185	232
Flood-prone policies (%)	53.3	23.8	37.8	55.2	70.8

Table 2: Insurance Demand and Risk Maps Digitalization

This table shows results from the two-way fixed effect regression: $Y_{it} = \alpha_i + \lambda_t + \beta * Digitalization_{it} + X_{it} + \epsilon_{it}$. The panel covers 3,053 counties from January 2010 to August 2019. The dependent variable Y_{it} measures the number of flood insurance policies in-force (with January 2010 normalized to 100) in county i in month t. The main explanatory variable $Digitalization_{it}$ is a binary variable indicating if county i has published the digitalized maps at time t; this term is set to zero for the control counties without digitalization. $Digitalization_{it}$ aggregates the community-level digitalization process to the county level: treatment is defined as more than 50 percent of the communities in county i publish the digitalized maps in the same month. Alternatively, $Digitalization_{it}$ is defined as a continuous variable that equal to the cumulative fraction of digitalized communities in county i in month t. α_i and λ_t are the county and year-month fixed effects. X_{it} are the covariates: $Premium_{it}$ is the average premium per policy (in \$), and $Coverage_{it}$ is the average coverage per policy (in \$k). Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)
Digitalization	21.39***	21.41***	20.57***	24.97***
	(5.66)	(5.66)	(5.13)	(5.78)
Premium		-0.003	-0.005	
		(0.01)	(0.01)	
Coverage		-0.069	0.005	
		(0.05)	(0.10)	
Observations	347,852	347,852	176,251	347,852
R-squared	0.69	0.69	0.75	0.69
Include never-treated	Y	Y	N	Y
Treatment construction	Discrete	Discrete	Discrete	Continuous

Table 3: House Prices and Risk Maps Digitalization

This table shows results from the two-way fixed effect regression: $Y_{it} = \alpha_i + \lambda_t + \beta * Digitalization_{it} + X_{it} + \epsilon_{it}$. The dependent variable Y_{it} measures the house price index of county i in month t. The house price data is obtained from Zillow. Zillow provides separate county-level house price indices for different types of houses, such as All Homes, Single-Family Homes, Top-tier Homes (within the 65th to 95th percentile range for a given county), Bottom-tier Homes (within the 5th to 35th percentile range for a given county), and Homes with 1, 2, 3, 4 or 5+ bedrooms. The main explanatory variable $Digitalization_{it}$ is an indicator variable which turns on if county i has launched digitalized risk maps at time t; this term is set to zero for the control counties without events. α_i and λ_t are the county and time fixed effects. X_{it} are the covariates. *Income* is the median household income (in $1,000); *Unemployment* is the unemployment rate (in percent). Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)
	Panel A: Zillow Home Value Index with Various Price Ranges				
Digitalization	-1.86***	-1.69***	-1.66***	-1.56***	-1.99***
	(0.46)	(0.42)	(0.43)	(0.38)	(0.45)
Observations	312,901	289,881	289,851	289,994	289,623
R-squared	0.80	0.79	0.79	0.79	0.79
House Type	All Homes	All Homes	Single-Family	Top Tier	Bottom Tier
Covariates	N	Y	Y	Y	Y
	Panel B: Zillow Home Value Index with Various Sizes				
Digitalization	-2.25***	-2.23***	-1.64***	-1.45***	-1.28***
	(0.52)	(0.45)	(0.42)	(0.41)	(0.42)
Observations	256,995	287,065	289,833	288,203	281,498
R-squared	0.72	0.75	0.79	0.80	0.79
House Type	1 bedroom	2 bedrooms	3 bedrooms	4 bedrooms	5+ bedrooms
Covariates	Y	Y	Y	Y	Y

Table 4: Salience and Past Flood Occurrence

This table shows results from the two-way fixed effect regression: $Y_{it} = \alpha_i + \lambda_t + \beta * Digitalization_{it} + X_{it} + \epsilon_{it}$. All the variables are defined as per Table 2. Panel A runs the regression in subsamples of counties that have not had any flood in the previous n years prior to the publication of the digitalized maps. Panel B uses subsamples of counties that have had at least one flood in the previous n years before digitalization. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
	Panel A: no flood in the past n years					
Digitalization	22.37***	25.74***	22.94**	28.08**	35.14**	47.79**
	(6.86)	(8.79)	(9.44)	(12.08)	(15.62)	(21.17)
Observations	125,560	93,121	66,571	47,745	30,718	19,358
R-squared	0.72	0.71	0.72	0.71	0.71	0.69
	Panel B: with a flood in the past n years					
Digitalization	14.45***	13.84***	19.13***	17.24***	16.11***	16.09***
	(4.85)	(4.60)	(6.30)	(5.42)	(4.93)	(4.60)
Observations	48,104	80,543	107,093	125,919	142,946	154,306
R-squared	0.83	0.79	0.77	0.77	0.77	0.77

Table 5: Heterogeneity of the Digitalization Effect Across Counties

This table shows results from the regression: $Y_{it} = \alpha_i + \lambda_t + \beta * Digitalization_{it} + \gamma * Digitalization_{it} \times Z_i + \epsilon_{it}$. Z_{it} captures the heterogeneity of ex-ante awareness of flood risk across counties. $Income_i$ is the median household income (in \$1,000) in county i. $Education_i$ is the percentage of people with college degrees (in percent). $ClimateOpinion_i$ is the percentage of people (in percent) who answered "Yes" to the question of whether they think global warming will harm them personally, which is obtained from the Yale climate opinion survey (Howe et al., 2015). $\mathbb{1}(Coastal)_i$ is a binary variable indicating if county i is from a coastal state or not. $\mathbb{1}(HighRisk)_i$ is a binary variable indicating if county i's proportion of the Special Flood Hazard Area (SFHA) is above the nationwide median. The SFHA defined by the NFIP as an area with a 1-percent or higher probability of being inundated in any given year. $RiskLevel_i$ is a continuous proxy for county i's flood risk level, which is the proportion of SFHA (in percent). $RiskLevel_i^2$ is the square of $RiskLevel_i$. All the other variables are defined as per Table 2. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Digitalization	50.55***	35.42***	81.11***	112.61***	45.96***	135.49***	21.46
	(13.97)	(9.41)	(24.85)	(34.08)	(12.38)	(38.94)	(21.44)
Digitalization ×							
Income	-0.65**			-0.59**		-0.78***	
	(0.27)			(0.27)		(0.29)	
Education		-0.65**					
		(0.31)					
ClimateOpinion			-1.56**	-1.72**		-1.54**	
			(0.63)	(0.67)		(0.71)	
$\mathbb{1}(Coastal)$					-21.61**	-19.73**	
					(9.09)	(9.18)	
$\mathbb{1}(HighRisk)$					-26.96***	-29.76***	
					(10.21)	(10.53)	
RiskLevel							1.02
							(0.83)
$RiskLevel^2$							-0.017**
							(0.0079)

Table 6: Flood Insurance Purchases and Social Connectedness

This table shows results from the event study: $Y_{it}^f = \beta_0 + \beta_1 * Connected_i^f \times Post_t^f + \beta_2 * Connected_i^f + \beta_3 * Post_t^f + \epsilon_{it}^f$. For a given flood event f and the associated flooding counties $\{j\}_f$, county i's social connectedness to $\{j\}_f$ is measured by the relative probability of Facebook friendship $p_{i,f} = \sum_{\{j\}_f} w_j * p_{i,j}$, where $p_{i,j}$ is the county-by-county probabilities obtained from Bailey et al. (2018b). w_j represents population-weighting or equal-weighting scheme. $Connected_i$ is a binary variable indicating if county i is socially connected with the flooding area, which is defined as having a value of $p_{i,f}$ above the state-median. The analysis sample consists of only counties that are geographically distant to the flooding area. Three different choices of distance threshold are considered: being 500, 750 and 1,000 miles. $Post_t$ is a binary variable indicating post-flood periods. The dependent variable Y_{it} measures the insurance demand in county i in month t, which is defined as per Table 2. Panel A uses the full sample period from January 2010 to August 2019. Panel B uses a restricted sample period from January 2014 to December 2017. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)
	Panel A: Full Sample Period 2010-2019			
$Connected * Post$	1.11***	0.90***	0.86***	1.54***
	(0.20)	(0.20)	(0.19)	(0.24)
Observations	23,369,392	23,369,392	32,114,276	16,335,272
	Panel B: Restricted Sample Period 2014-2017			
$Connected * Post$	1.28***	0.86***	0.77***	1.72***
	(0.28)	(0.26)	(0.23)	(0.35)
Observations	8,760,517	8,760,517	12,028,168	6,019,329
Connectedness Weight	PW	EW	PW	PW
Distance Threshold	750	750	500	1000

Table 7: Heterogeneity of Social Connectedness and Flood Salience

This table shows results from the event study: $Y_{it}^f = \beta_0 + \beta_1 * Connected_i^f \times Post_t^f + \beta_2 * Connected_i^f + \beta_3 * Post_t^f + \epsilon_{it}^f$. Panel A and B consider two deviations from the baseline specifications presented in Table 6; otherwise, the variables are defined as per Table 6. In Panel A, $Connected_i$ is defined as a binary variable that equals one if county i has a connectedness measure in the top quartile of the state and equals zero if county i has a connectedness measure in the bottom quartile of the state. Panel B only includes the significant flood events defined by the FEMA. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)
	Panel A: Top vs. Bottom Quartiles of Connectedness			
$Connected * Post$	2.20***	1.74***	1.91***	2.98***
	(0.30)	(0.30)	(0.30)	(0.37)
Observations	11,965,523	11,965,523	12,576,973	8,101,031
	Panel B: Subsample of Significant Floods			
$Connected * Post$	3.08**	4.57***	4.12**	5.95***
	(1.43)	(1.44)	(1.64)	(2.02)
Observations	688,352	688,352	1,107,478	381,157
Connectedness Weight	PW	EW	PW	PW
Distance Threshold	750	750	500	1000

Table 8: Alternative Methodology of Estimating the Causal Effect of Social Connectedness

This table shows results from regression: $log(Policies)_{i,t} = \beta * FriendFlood^{Distant}_{i,t-k,t} + FE_{state\times time} + \epsilon_t$. Following the methodology proposed by Bailey et al. (2018a), $FriendFlood^{N}_{i,t-k,t}$ measures the average flood experience of a county i's social network N between $t-k$ and t. $FriendFlood^{N}_{i,t-k,t}$ is calculated as the weighted average as $\sum_j \theta^N_{i,j} * Flood_{j,t-k,t}$, where $\theta^N_{i,j}$ is share of county i's friends in network N who live in county j, and $Flood_{j,t-k,t}$ is the number of floods in county j between $t-k$ and t. A geographically distant network $N = Distant$ is a set of counties that are certain miles away from county i. Columns 1 through 4 show results of using 750 miles as the threshold; columns 5 through 7 use 250, 500, and 1,000 miles, respectively. The measurement window (i.e. k) of floods takes values of 3, 6, 12 or 24 months. $FE_{state\times time}$ are the state×time fixed effects. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Geographically Distant Friends' Flood Experiences						
$FriendFlood^{Distant}_{i,t-k,t}$	0.013***	0.012***	0.011***	0.013***	0.009***	0.011***	0.012***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Observations	321,511	349,443	340,120	284,729	321,511	321,511	321,511
R-squared	0.42	0.42	0.42	0.42	0.42	0.42	0.42
	Panel B: Geographically Distant Friends' Significant Flood Experiences Only						
$FriendFlood^{Distant}_{i,t-k,t}$	0.042***	0.041***	0.041***	0.033***	0.043***	0.067***	0.042***
	(0.004)	(0.005)	(0.004)	(0.003)	(0.005)	(0.006)	(0.004)
Observations	321,511	349,443	340,120	284,729	321,511	321,511	321,511
R-squared	0.42	0.42	0.42	0.42	0.42	0.42	0.42
Distance Threshold	750 miles	750 miles	750 miles	750 miles	250 miles	500 miles	1,000 miles
Flood Window	12 months	3 months	6 months	24 months	12 months	12 months	12 months

Appendix

A. Multiple Publications of Risk Maps

As discussed in Section 2.3, the FEMA's Community Status Information only provides the publication date of the latest map. As the FEMA aims to review their maps every five years, in principle, a county may have two publication dates during my 10-year sample period, and in which case, the first publication should capture the digitalization of interest. In this section, I present a set of evidence to show that the FEMA fails this goal, and in reality, new publications take place much longer than every five years.

First, in an official audit report titled "FEMA Needs to Improve Management of Its Flood Mapping Programs" published in September 2017, evidence suggests that more than half of the database falls behind schedules.

Second, according to the FEMA's Community Status Information (as of June 2020), almost 75% of the communities have an effective date more than five years old, i.e., the latest update was before June 2015. Moreover, 37% or 13% of the maps are more than 10 or 20 years old. These statistics indicate that the FEMA has been struggling to keep pace with its goal.

Third, I have downloaded the Community Status Information at two points in time—December 2019 and June 2020. By comparing the effective dates in the two downloads, I can identify a sample of communities that have published new maps in 2020, i.e., the communities with two different dates in the two downloads. For these communities, I can impute the time spell between the two publications. I find 412 such cases in total, and on average, it takes 11.5 years to publish a new map.

B. The New Map Service Center and Information Cost

In July 2014, the FEMA launched a new online portal, known as the Map Service Center (MSC), to replace the legacy one. The new portal enhances address search, integrates products, improves user interface, and provides a variety of other upgrades and new features. More details can be found in the FEMA's newsletter.[22]

Appendix Figure A.9 presents screenshots of the new and old websites, obtained from the Wayback Machine. Consistent with the timeline discussed above, the old portal's last appearance was on July 22, 2014, and the new MSC is available since July 28, 2014. As shown, the announcement on the website said, "Welcome to the New FEMA Flood Map Service Center! A series of major changes, including a complete site redesign, have taken effect on the MSC. All flood hazard products are now available free of charge, and the former products catalog has been replaced with an integrated Search All Products feature that allows you to find and download all products for a geographic area."

In this setting, I construct the treatment group as counties that had published the digitalized maps before the new MSC. Thus, households in the treated counties have experienced using both portals. In comparison, the control group consists of counties with no digitalized maps yet. Thus, the upgrade is irrelevant. I estimate a standard difference-in-difference model:

$$Y_{it} = \beta_0 + \beta_1 * Treated_i \times Post_t + \beta_2 * Treated_i + \beta_3 * Post_t + \epsilon_{it} \tag{7}$$

$Treated_i$ is the treatment dummy indicating whether county i has digitalized risk maps. $Post_t$ is a binary variable indicating if t is posterior to July 2014. β_1 is the difference-in-differences estimate of interest.

[22] https://www.fema.gov/media-library-data/1405342400259-4b9d70489f7e9f6ffd90ba001182f112/The+New+FEMA+Flood+Map+Service+Center.pdf

Appendix Figures and Tables

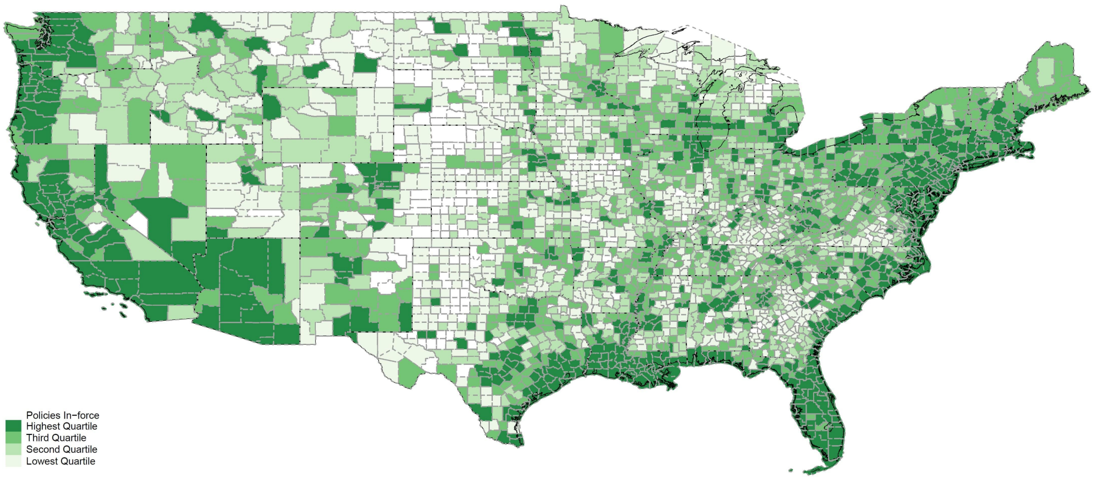


Figure A.1. The Number of Policies In-force (Averages of 2010-2019)

This figure shows a heat map of the geographical distribution of the number of flood insurance policies in-force at the county level. Darker shades represent higher densities.

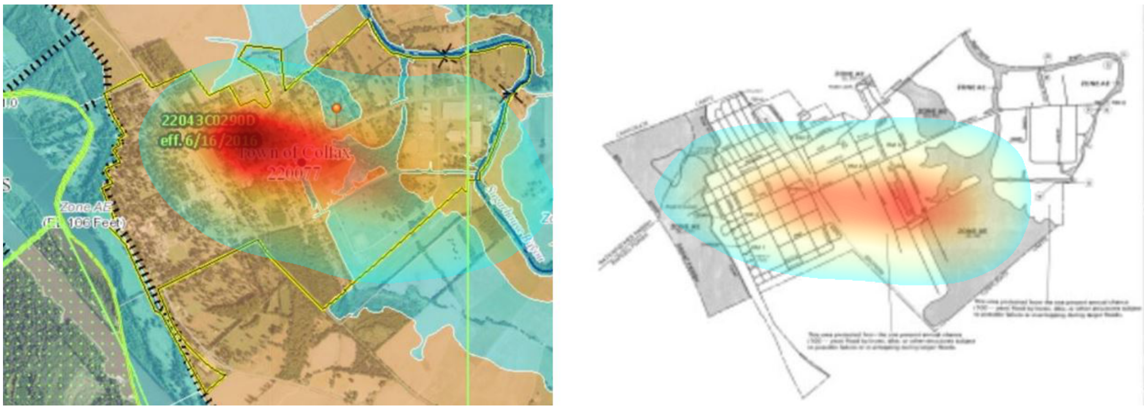


Figure A.2. Predicted Human Eye Fixations

This figure shows the digitalized flood hazard map and the corresponding scanned paper map side by side as one image, and it shows an overlay heat map of predicted human eye fixations on the image. The prediction is generated by a machine-learning-based methodology called the Saliency Attentive Model (SAM) developed by Cornia et al. (2018). The darker the heat map, the more attention is allocated to that area of the image.

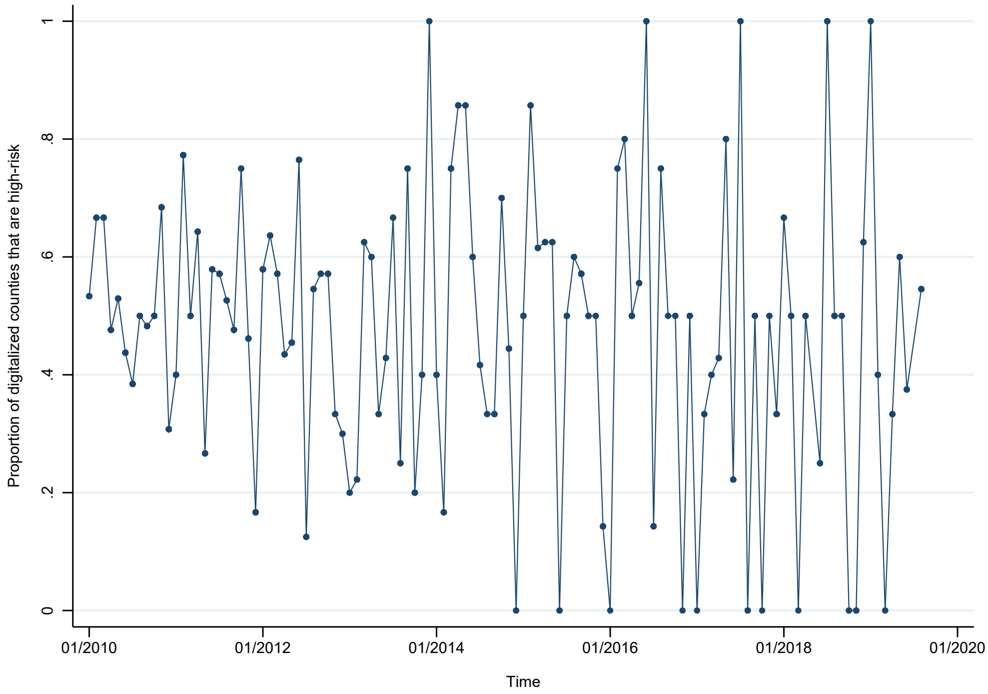


Figure A.3. Digitalization Timing and Flood Risk Levels

This figure plots the time-series of the share of high-risk counties among newly digitalized counties. For each month, the y-variable is calculated as the number of high-risk counties that publish digitalized maps divided by the total number of counties that publish digitalized maps in that month. A county is defined as high-risk if its proportion of the Special Flood Hazard Area (SFHA) is above the nationwide median. The SFHA defined by the NFIP as an area with a 1-percent or higher probability of being inundated in any given year.

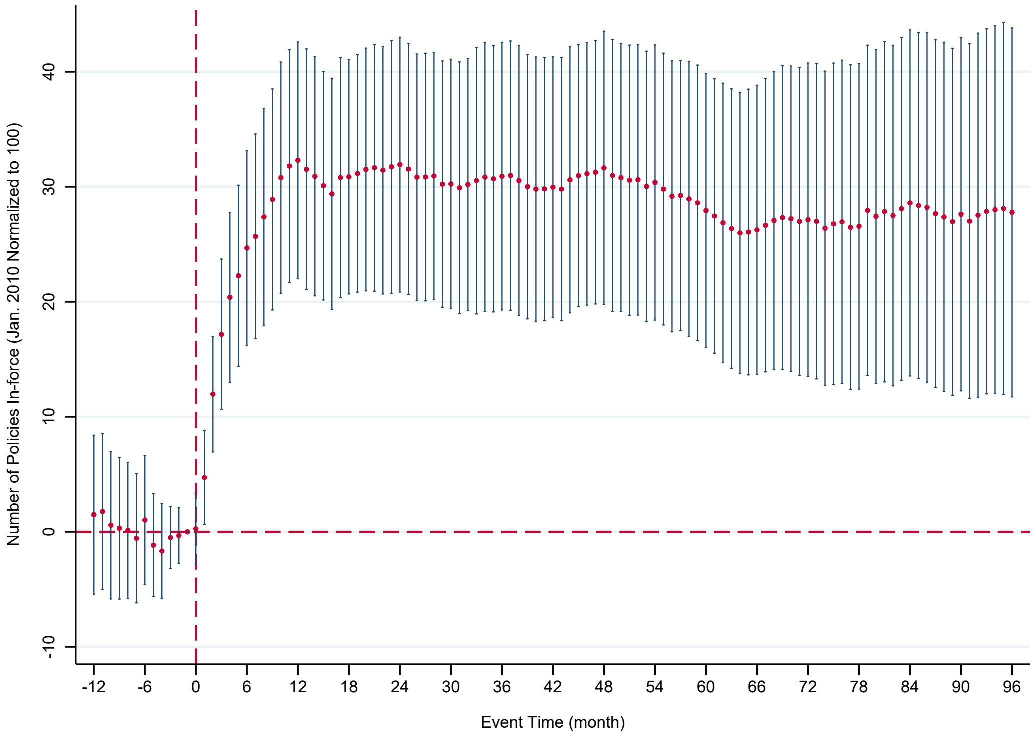


Figure A.4. The Impact of Risk Map Digitalization: Normalized Outstanding Policies

This figure shows the dynamic effects of risk maps digitalization on insurance purchases. It plots the coefficient estimates of $\{\beta_k\}$ in the regression: $Y_{it} = \alpha_i + \lambda_t + \sum_k \beta_k * \mathbb{1}(t = t_i^* + k) + \epsilon_{it}$. $\{\beta_k\}$ are measured relative to β_{-1} which is omitted. The dependent variable Y_{it} measures the number of flood insurance policies in-force (with January 2010 normalized to 100) in county i in month t. t_i^* is the publication time of the digitalized maps in county i. $\mathbb{1}(t = t_i^* + k)$ is set to zero for the untreated. α_i and λ_t are the county and year-month fixed effects. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.

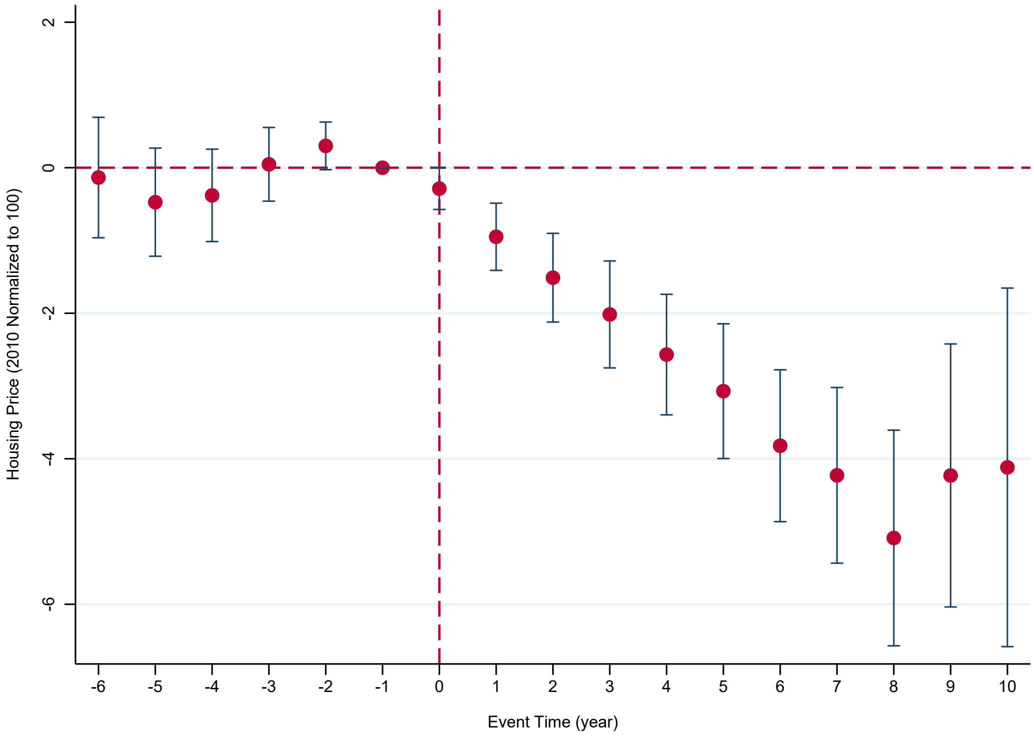
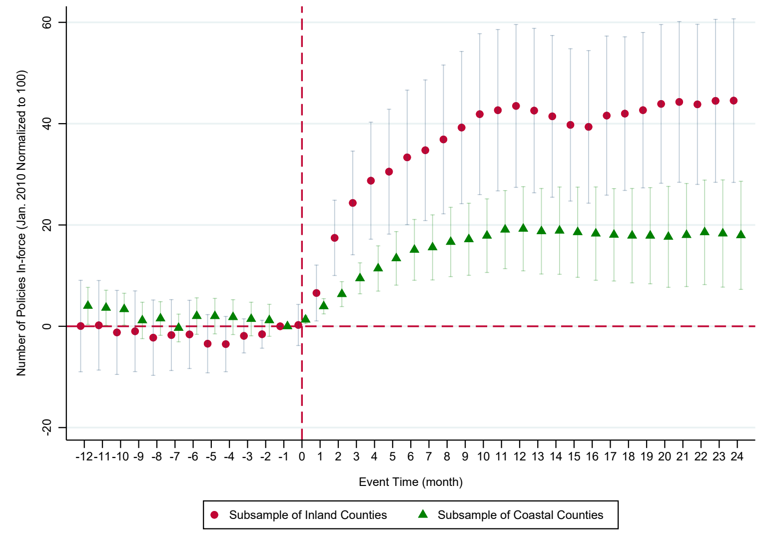
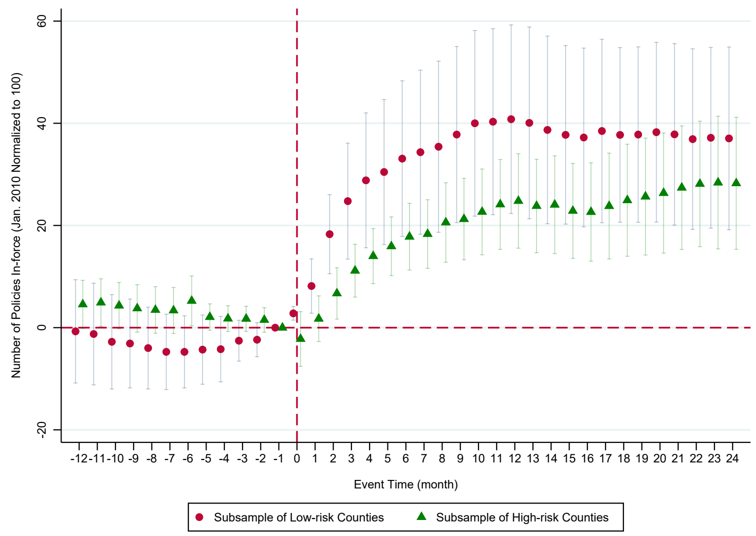


Figure A.5. The Impact of Risk Maps Digitalization on House Prices

This figure shows the dynamic effects of risk maps digitalization on county-level house price index. It plots the coefficient estimates of $\{\beta_k\}$ in the regression: $HousePrice_{it} = \alpha_i + \lambda_t + \sum_k \beta_k * \mathbb{1}(t = t_i^* + k) + \epsilon_{it}$. $\{\beta_k\}$ are measured relative to β_{-1} which is omitted. The dependent variable $HousePrice_{it}$ is the house price index (with January 2010 normalized to 100) in county i in year t, which is obtained from the Federal Housing Finance Agency. t_i^* is the year when county i publishes its digitalized maps. $\mathbb{1}(t = t_i^* + k)$ is set to zero for the untreated. α_i and λ_t are the county and year fixed effects. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.



(a) Subsamples of Inland vs. Coastal Counties



(b) Subsamples of High vs. Low Share of SFHA

Figure A.6. Heterogeneity by Flood Risk Across Counties

This figure shows the dynamic effects of risk maps digitalization on insurance purchases, in subsamples. All the variables and the regression are defined as per Figure 3. Figure (a) splits the sample by whether the county is from a coastal state or not. Figure (b) splits the sample by whether the county has an above- or below-median value of the proportion of the Special Flood Hazard Area (SFHA). The SFHA defined by the NFIP as an area with a 1-percent or higher probability of being inundated in any given year. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.

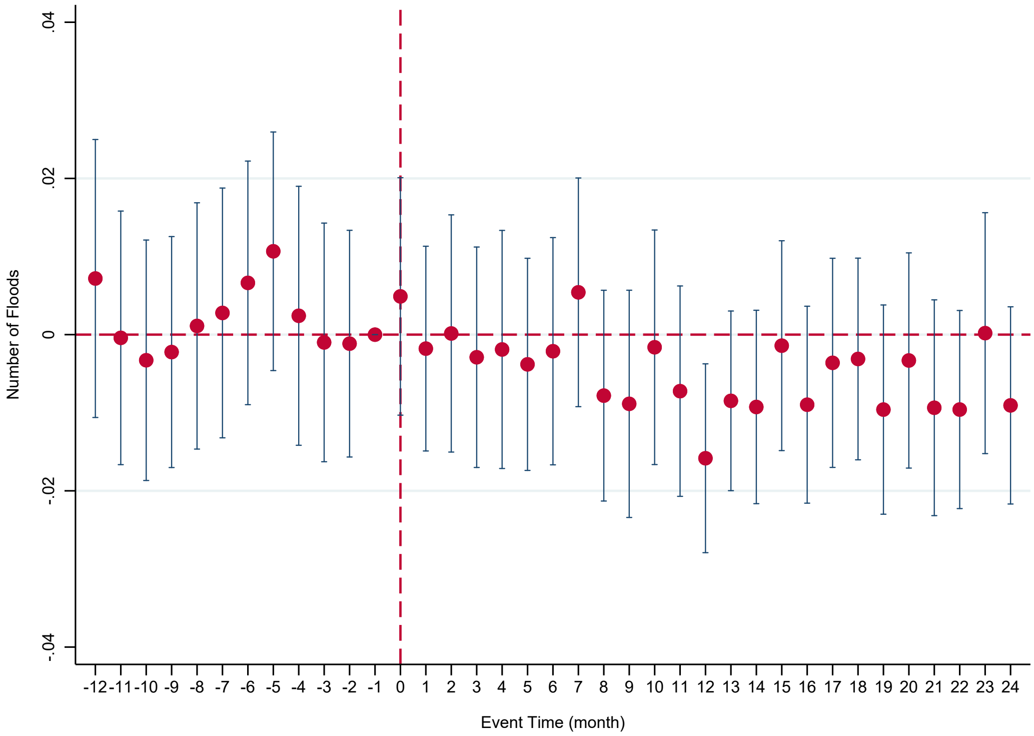
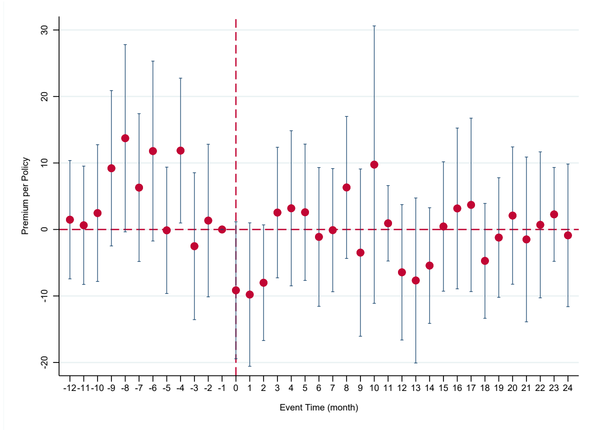
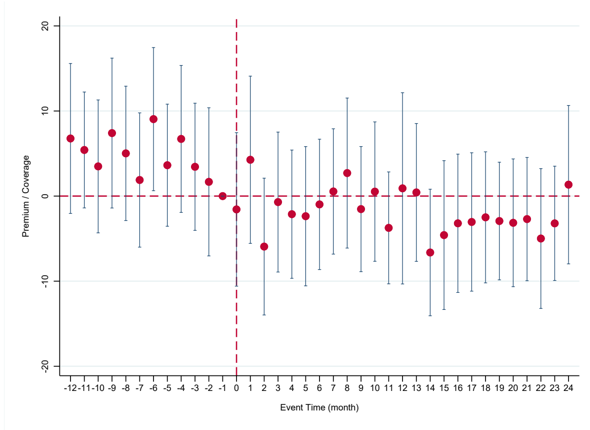


Figure A.7. The Impact of Digitalization on Flood Occurrence

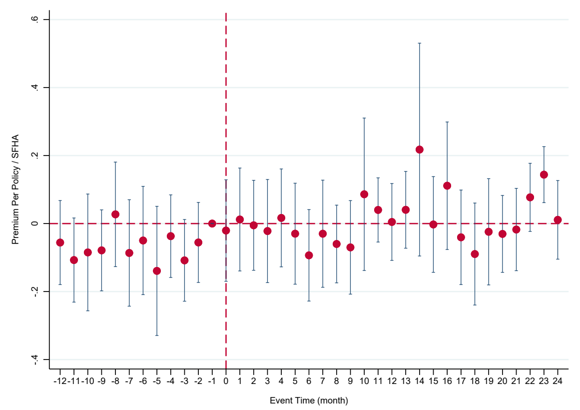
This figure shows the dynamic effects of risk maps digitalization on flood occurrence. It plots the coefficient estimates of $\{\beta_k\}$ in the regression: $Y_{it} = \alpha_i + \lambda_t + \sum_k \beta_k * \mathbb{1}(t = t_i^* + k) + \epsilon_{it}$. The dependent variable Y_{it} measures the number of floods in county i in month t. All the other variables are defined as per Figure 3 (or Figure A.4). Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.



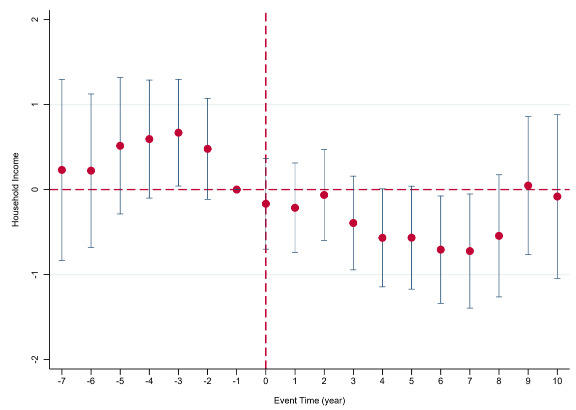
(a) Premium



(b) Premium/Coverage



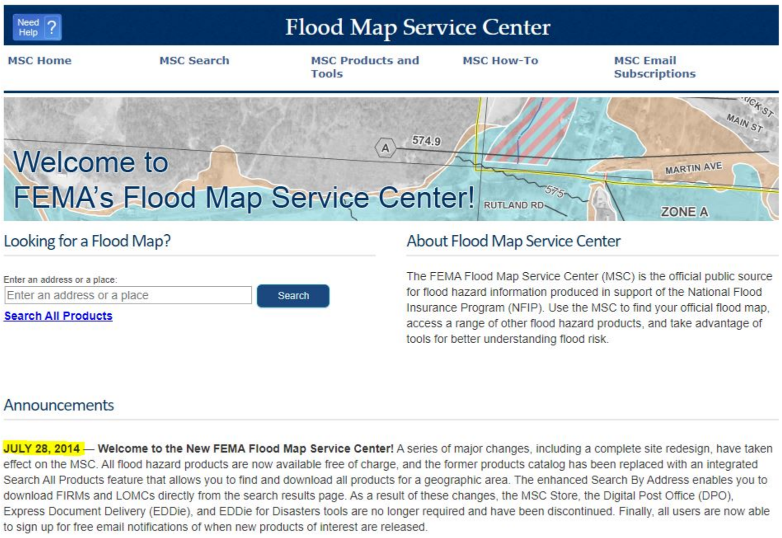
(c) Premium/SFHA



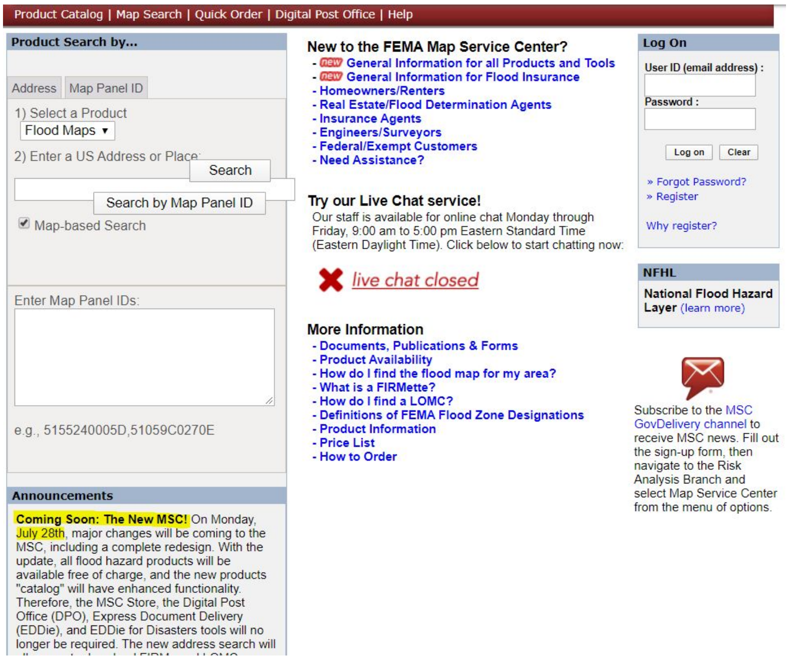
(d) Household Income

Figure A.8. Digitalization and The Price of Flood Insurance

This figure shows the dynamic effects of risk maps digitalization on the cost of buying a flood insurance policy. It plots the coefficient estimates of $\{\beta_k\}$ in the regression: $Y_{it} = \alpha_i + \lambda_t + \sum_k \beta_k * \mathbb{1}(t = t_i^* + k) + \epsilon_{it}$. In figure (a), Y_{it} measures the average premium per policy in county i in month t. In figure (b), Y_{it} measures the average premium per policy per \$1000 coverage in county i in month t. In figure (c), Y_{it} measures the average premium per policy divided by the fraction of SFHA in county i in month t. In figure (d), Y_{it} measures the median household income in county i in year t. All the other variables are defined as per Figure 3 (or Figure A.4). Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.



(a) New Portal of the Map Service Center



(b) Legacy Portal of the Map Service Center

Figure A.9. The New FEMA Map Service Center (MSC) Launched in July 2014

This figure shows the FEMA's new and old portal of its online GIS database. The screenshots were taken on different dates in July 2014 by the Wayback Machine. The new Map Service Center (MSC) was officially launched on 28 July, 2014. The last appearance of the old website (in the library of the Wayback Machine) was on 22 July, 2014.

Table A.1: Insurance Demand and Risk Maps Digitalization

This table shows results from the two-way fixed effect regression: $Y_{it} = \alpha_i + \lambda_t + \beta * Digitalization_{it} + X_{it} + \epsilon_{it}$. The main explanatory variable $Digitalization_{it}$ is a binary variable indicating if county i has published the digitalized maps at time t; this term is set to zero for the control counties without digitalization. $Digitalization_{it}$ aggregates the community-level digitalization process to the county level: treatment is defined as more than 50, 75 or 90 percent of the population in county i obtain access to the digitalized maps in the same month. All the other variables are defined as per Table 2. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)
Digitalization	20.89***	20.87***	18.63***	25.25***	26.28***
	(6.78)	(6.66)	(6.23)	(6.61)	(7.03)
Premium		-0.003	-0.023*		
		(0.007)	(0.012)		
Coverage		-0.070	-0.024		
		(0.050)	(0.097)		
Observations	347,852	347,852	241,271	347,852	347,852
R-squared	0.69	0.69	0.70	0.69	0.69
Include never-treated	Y	Y	N	Y	Y
Treatment definition	>50%	>50%	>50%	>75%	>90%

Table A.2: Various Specifications of Event Windows

This table shows results from the two-way fixed effect regression: $Y_{it} = \alpha_i + \lambda_t + \beta * Digitalization_{it} + X_{it} + \epsilon_{it}$. All variables are defined as per Table 2. In this table, each specification considers a different sample choice, where only observations within a certain leads and lags around the digitalization date are included. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Digitalization	20.37***	20.50***	20.55***	20.02***	21.52***	21.63***	21.59***	20.93***
	(5.63)	(5.53)	(5.42)	(5.24)	(5.73)	(5.63)	(5.51)	(5.30)
Observations	289,251	274,681	259,459	244,035	284,868	270,298	255,076	239,652
R-squared	0.66	0.65	0.64	0.63	0.66	0.65	0.64	0.63
Leads	60	60	60	60	48	48	48	48
Lags	60	48	36	24	60	48	36	24
Digitalization	23.19***	23.24***	23.04***	22.09***	24.32***	24.41***	24.10***	22.93***
	(5.99)	(5.87)	(5.73)	(5.47)	(5.90)	(5.82)	(5.71)	(5.49)
Observations	279,583	265,013	249,791	234,367	273,062	258,492	243,270	227,846
R-squared	0.66	0.65	0.64	0.63	0.67	0.66	0.65	0.64
Leads	36	36	36	36	24	24	24	24
Lags	60	48	36	24	60	48	36	24

Table A.3: Salience and Past Floods

This table repeats the analysis in Table 4 with an extra restriction on the sample selections. This table shows results from the two-way fixed effect regression: $Y_{it} = \alpha_i + \lambda_t + \beta * Digitalization_{it} + X_{it} + \epsilon_{it}$. All the variables are defined as per Table 4. The subsamples of counties are constructed such that: (1) they have not had any flood in the previous n years prior to the publication of the digitalized maps; (2) they have low inherent flood risk, which is defined as having a below-median value of $HighRisk$. $HighRisk$ is defined as per Table 3. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)	(6)
	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
	low-risk counties & no flood in the past n years					
Digitalization	31.91***	40.31**	35.24**	40.41**	49.53*	66.49*
	(12.21)	(15.80)	(16.35)	(20.23)	(25.96)	(34.14)
Observations	66,359	49,761	37,265	26,884	17,745	11,598
R-squared	0.70	0.69	0.70	0.70	0.69	0.69

Table A.4: Insurance Purchases in SFHA and Non-SFHA

This table shows results from the two-way fixed effect regression: $Y_{it} = \alpha_i + \lambda_t + \beta * Digitalization_{it} + \epsilon_{it}$. The dependent variable Y_{it} differs in columns 1 through 4. The first specification examines the number of policies in-force held by SFHA households, with the value of January 2010 normalized to 100. The second specification examines the number of policies in-force held by non-SFHA households, with the value of January 2010 normalized to 100. The third specification takes the difference between the two. The fourth specification examines the fraction of policies held by SFHA households relative to the total policies in-force. All the other variables are defined as per Table 2. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)
Digitalization	7.96**	53.02***	-43.64**	-0.047***
	(3.34)	(20.11)	(20.46)	(0.005)
Constant	106.70***	112.85***	-5.17	0.55***
	(1.27)	(7.66)	(7.83)	(0.002)
Dependent Variable	SFHA Policies	Non-SFHA Policies	SFHA – Non-SFHA	SFHA/Total
Observations	329,114	339,042	320,484	347,672
R-squared	0.75	0.65	0.65	0.86

Table A.5: Upgrade of Map Service Center and Insurance Purchase

This table shows results from the difference-in-differences regression: $Y_{it} = \beta_0 + \beta_1 * Treated_i \times Post_t + \beta_2 * Treated_i + \beta_3 * Post_t + \epsilon_{it}$. $Treated_i$ is the treatment dummy indicating whether county i has digitalized risk maps or not. $Post_t$ is a binary variable indicating if t is posterior to July 2014, which is the launch date of the FEMA's new online portal of digital maps (called the Map Service Center). The other variables are defined as per Table 2. β_1 is the difference-in-differences estimate of interest. I run the regression in a sample from July 2011 to July 2017, i.e. three years before and after the launch of the new portal. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)	(6)
$Treated \times Post$	4.52	4.85	6.07	6.14	5.98	6.14
	(3.15)	(3.16)	(4.04)	(4.07)	(4.61)	(4.68)
Observations	216,401	216,401	216,401	216,401	216,401	216,401
R-squared	0.006	0.009	0.011	0.014	0.012	0.015
Covariates	N	Y	N	Y	N	Y
Digitalization Definition	>50%	>50%	>75%	>75%	>90%	>90%

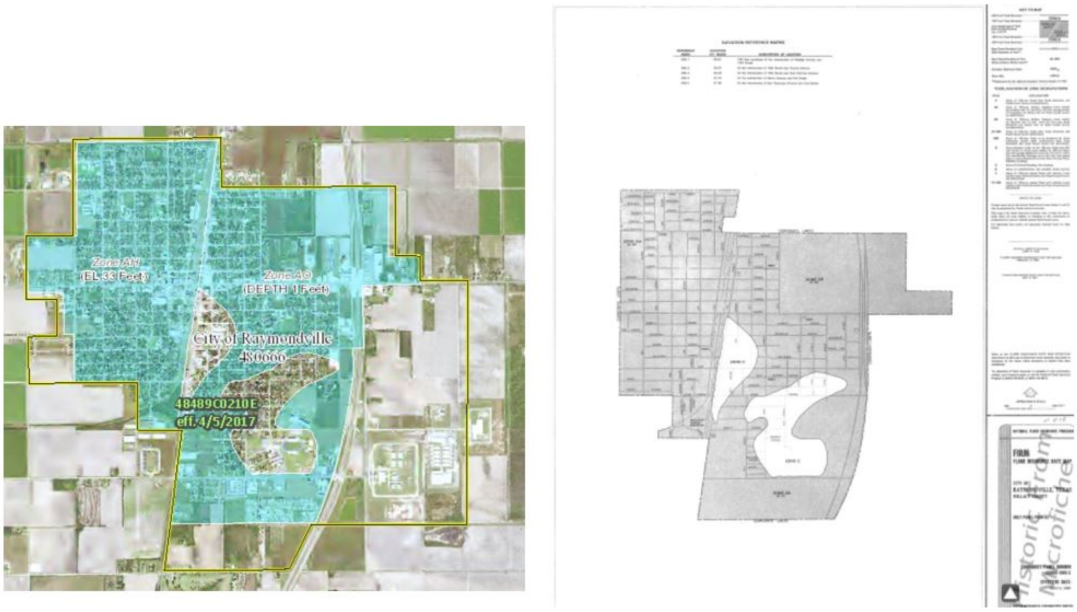
Table A.6: Past Floods Predicting Maps Digitalization

This table shows results from the regression: $Digitalization_{it} = \sum_k \beta_k * Flood_{i,t-k} + \epsilon_{it}$. $Digitalization_{it}$ is a binary variable indicating county i digitalizes flood risk maps in month t. $Flood_{i,t-k}$ is a binary variable indicating county i has a flood in time $t-k$. Standard errors are clustered at the county level and presented in parentheses. *** p<0.01, ** p<0.05, * p<0.1

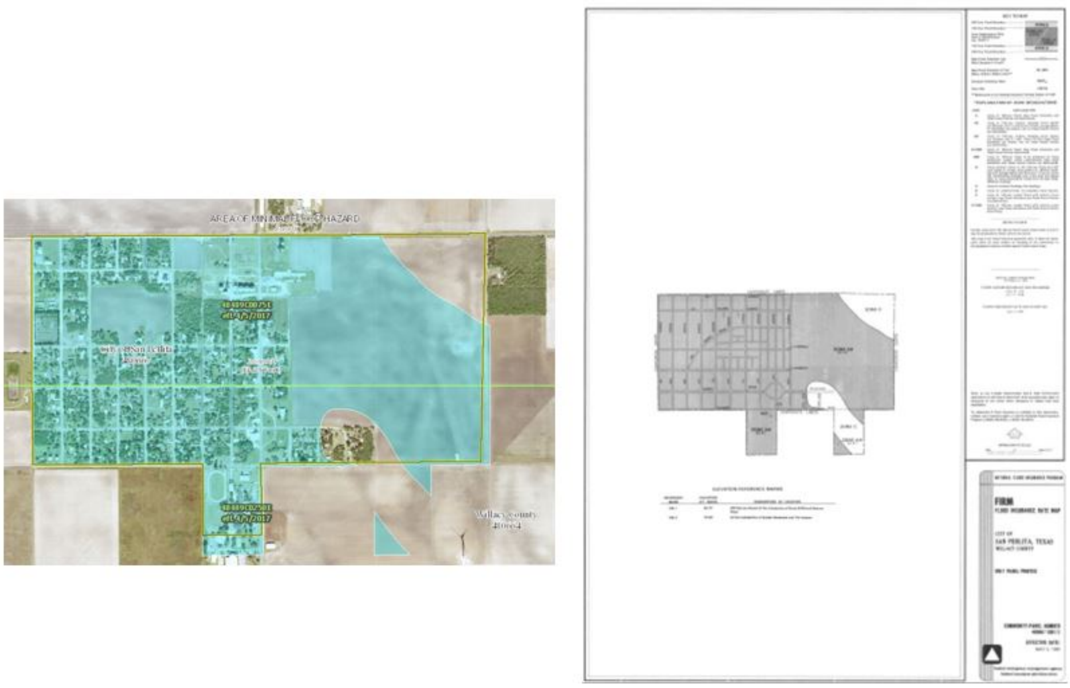
	(1)	(2)	(3)	(4)	(5)
$Flood_t$	0.001		0.001	0.001	0.001
	(0.001)		(0.001)	(0.001)	(0.001)
$Flood_{t-1}$		-0.001	-0.001	-0.001	-0.001
		(0.001)	(0.001)	(0.001)	(0.001)
$Flood_{t-2}$				-0.000	-0.000
				(0.001)	(0.001)
$Flood_{t-3}$				0.000	0.000
				(0.001)	(0.001)
$Flood_{t-4}$					0.000
					(0.001)
$Flood_{t-5}$					0.001
					(0.001)
$Flood_{t-6}$					0.000
					(0.001)
$Constant$	0.003***	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	359,740	356,611	356,611	350,361	341,014
R-squared	0.00	0.00	0.00	0.00	0.00

Web Appendix

* Due to size limit, click here to view the full appendix online.



Example 1: City of Raymondville, Willacy County, Texas



Example 2: City of San Perlita, Willacy County, Texas