Social Interactions and Households' Flood Insurance Decisions

Zhongchen Hu*

Chinese University of Hong Kong (Shenzhen)

January 2022

Abstract

Flooding is the most costly natural disaster faced by US households, yet policymakers are puzzled by the low take-up rates for flood insurance. Leveraging novel transaction-level data, this paper studies the influence of social interactions on households' insurance decisions. I show that households increase flood insurance purchases by 1–5 percent when their geographically distant friends are exposed to flooding events or to campaigns for flood insurance. These exogenous shocks to far-away friends should not affect local households' own insurance decisions except through peer effects. I provide evidence suggesting that social interactions facilitate learning through information dissemination and attention triggering.

*I am thankful for the advice of Ashwini Agrawal, Asaf Bernstein (discussant), Tom Chang (discussant), Fabrizio Core, Andreas Fagereng, Francisco Gomes, Juanita González-Uribe, Daniel Gottlieb, Benjamin Guin (discussant), Isaac Hacamo, Bob Hartwig, Dirk Jenter, Ankit Kalda, Peter Koudijs, Dong Lou, Diogo Mendes, Greg Niehaus, Daniel Paravisini, Cameron Peng, Paolo Sodini, Jan Starmans, Per Stromberg, and Su Wang. I am also grateful to seminar and conference participants at the London School of Economics, University of South Carolina, Chinese University of Hong Kong, Chinese University of Hong Kong (Shenzhen), University of Toronto, Stockholm School of Economics, BI Norwegian Business School, SGF Conference 2021, SEHO 2021 Annual Meeting, SFS Cavalcade North America 2021, and WFA 2021, for helpful comments. This paper was previously circulated under the title "Salience and Households' Flood Insurance Decisions." Address for correspondence: huzhongchen@cuhk.edu.cn

1 Introduction

Flooding, the most costly natural disaster in the US, is drawing increasing attention in public policy debates.[1] Flood risk materializes with low frequency but has catastrophic consequences for households' asset values and welfare, making insurance crucial to hedge this tail risk. However, US households' take-up rate of flood insurance is surprisingly low, even though the government has spent $36.5 billion on subsidies to encourage it.[2]

Understanding the factors that influence households' insurance decisions is thus important from a policy perspective to resolve the underutilization of flood insurance. In this paper, I examine the effect of social interactions in insurance markets, in light of the recent theoretical work and empirical evidence suggesting that social interactions shape economic outcomes (Shiller, 2017, 2020; Hirshleifer, 2020). As a comparison, in classical models of economics and finance, agents make decisions in a social vacuum and interact only impersonally via market prices.

Identifying a causal effect of social interactions on insurance demand faces several challenges. First, peer groups are formed endogenously and are subject to common shocks (Manski, 1993). For example, in the aftermath of a flood, an individual and her local friends may all decide to purchase flood insurance without influencing each other. Second, 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. Third, the determinants of insurance demand—such as actual risk, materialized flood experiences, attention, risk aversion, and social interactions—often act simultaneously, making it difficult to isolate the effect of social interactions.

I overcome these challenges by using data from the National Flood Insurance Program (NFIP) and using two identification strategies. The NFIP was created by the US Congress in 1968 to provide affordable household flood insurance. I obtain over 50 million transaction-level observations from January 2009 to August 2019, including information

[1] According to the National Oceanic and Atmospheric Administration, in the 2010s, flood-related events caused losses of $658 billion, substantially outpacing other natural disasters.

[2] The Insurance Information Institute 2016 Survey suggests that only 12% of US households had flood insurance. Even in flood-prone areas, the take-up rate is as low as 30% (Kousky et al., 2018). Consistently, less than 20% of the homes flooded by Hurricanes Harvey and Sandy were insured. Given that flood insurance is mandatory for mortgages in the federally-designated special flood hazard areas (SFHAs), the already-low overall take-up rate suggests that most households do not purchase flood insurance when it is not required (i.e., outside the SFHAs or without mortgages). 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).

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 other property and casualty risks, flood risk has been shunned by private insurers (Horn and Webel, 2019), leaving few outside options for households.[3] 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 level, the change in the number of policies in-force should capture shifts in the demand curve.

To identify the effect of social interactions, I leverage an endogenously formed structure of social networks on Facebook and exploit random shocks to friends in the networks. Crucially, the shock should not affect an individual's own insurance decision in any way but through peer effects from friends who receive the shock. The social networks are captured by the county-pairwise Social Connectedness Index (Bailey et al., 2018b), which is based on anonymized information on the universe of friendship links between Facebook users in the US. Facebook's enormous scale and comprehensive market penetration make this measure a realistic representation of real-world US social networks.[4]

My first quasi-experiment exploits non-local flooding events as the shock to far-away friends. 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. Similarly, my second quasi-experiment exploits non-local campaigns for flood insurance take-up as the shock to geographically distant friends.

Pooling all major floods (that triggered federal assistance) between 2010 and 2019 in an event study design, my first experiment finds that the number of flood insurance policies in-force increases by 0.94% in counties that are more socially connected to the flooded area, compared to the less-connected counterfactuals in the same distant state. The ef-

[3]The private market for flood insurance barely exists because the government heavily subsidizes the NFIP (see footnote 11). Cutting subsidies and privatization are the subjects of contemporary policy debates, but irrelevant to this paper's research question.

[4]Facebook is the world's largest online social network, with over 234 million active users in the US and Canada and more than 1.9 billion users globally (Bailey et al., 2018a).

fect appears to be persistent; the incremental policies are mostly renewed in subsequent years. I also document two additional findings consistent with the hypothesis that social interactions affect households' insurance decisions. First, the effect is monotonic in the strength of social connectedness. Second, the most damaging floods cause the strongest effect transmitted across social networks; the effect is as large as 5.1% when I restrict my analysis to 18 significant floods, as defined by FEMA. The results are robust to various alternative specifications (such as different distance thresholds and sample periods) and to the alternative empirical approach proposed by Bailey et al. (2018a).

I evaluate two potential non-causal alternative explanations for my findings. First, one might be concerned that peers receive common shocks. However, my empirical design—by only comparing counties within the same state (e.g., California) far away from a given flooded area (e.g., Boston)—makes this explanation unlikely, as it requires that the more-connected-to-Boston California counties contemporaneously experience a flood, while the less-connected-to-Boston California counties do not. To provide further evidence, I show that there is no difference between the more- and less-connected counties in terms of flood occurrences (either in the past or ex-post) and inherent flood risk.

The second non-causal explanation could be that individuals who have experienced a flood may move to the connected counties (where families and friends live) and then purchase flood insurance for their own new homes, without necessarily affecting their peers.[5] However, I show that migration cannot explain my findings.

I then explore which mechanisms underlie the observed causal effect of social interactions (with far-away friends) on households' decisions to acquire flood insurance. I posit that social learning is an important channel through which households update their beliefs about flood risk and flood insurance. Interactions with peers can facilitate both direct and indirect social learning. Conceptually, for the *direct* channel, social interactions act as an informational source that disseminates pertinent and otherwise costly information; for the *indirect* channel, social interactions nudge attention to freely available but less salient public information, as households have limited cognitive resources and attention (Tversky and Kahneman, 1973, 1974; Brown et al., 2010; Lacetera et al., 2012; Bordalo et al., 2012, 2013; Hastings and Shapiro, 2013; Stango and Zinman, 2014; Busse et al.,

[5] A large literature shows 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).

2015; Andersen et al., 2020).

In my specific setting, direct learning probably plays a mild role and I provide some suggestive arguments. First, geographically distant floods do not contain any new information about local flood risk.[6] Second, unlike markets that are segmented and products that are differentiated (such as consumer loans), the US flood insurance market is subject to minimal informational frictions, since it is dominated by the government-backed, long-established NFIP offering standardized policies. Information about flood insurance and a given property's flood risk is fairly easy to acquire and process, either online or offline (see Section 2 for details). Third, I present survey evidence that after a flood, discussions on Facebook—as an example of social interactions in general—typically concern the disaster itself (such as pictures of the flood) than flood insurance. This result provides suggestive evidence that information dissemination about flood insurance seems second-order over online interactions.[7]

Social interactions can also facilitate social learning indirectly, even if the interactions may not transmit direct information pertaining to an individual's local flood risk or flood insurance per se. Prior studies (Chivers and Flores, 2002; Pope, 2008) provide suggestive evidence that households are inattentive to flood risk information.[8] Therefore, when far-away friends share flood experiences, one's attention may be drawn to flood risk, thereby causing learning from the public information set about flood insurance and/or one's own risk exposure. Such attention-triggered learning can be interpreted as a salience effect in the classical theoretical frameworks of salience. For example, in the model of Hirshleifer and Teoh (2003), information presented in salient form is, owing to limits to investor attention, absorbed more easily than information that is less salient or that is only implicit in the public information set. The persistent effect of social interactions documented in this paper is consistent with a learning process which leads to a permanent shift in insurance take-up.

[6] Gallagher (2014) and Dessaint and Matray (2017) show that the occurrence of a flood (or hurricane) does not even contain information about the probability of a future local flood, consistent with the climate literature (Elsner and Bossak, 2001; Pielke, 2005; Pielke et al., 2005; Landsea et al., 2006).

[7] However, it is important to note that the Facebook Social Connectedness Index is best viewed as a proxy for broader social links: connected individuals are likely to interact in general whether on or off Facebook. I thus acknowledge the limitation of the survey, as it does not observe broader social interactions (for example, friends may talk about the importance of flood insurance over the phone).

[8] For example, Chivers and Flores (2002) present survey evidence that the majority of households buying properties in FEMA-designated special flood hazard areas in Colorado are unaware of the degree of flood risk.

As an alternative mechanism, social interactions may influence an individual's insurance decision through her risk attitude. While my analysis does not allow me to measure changes in risk aversion, I present some evidence that my results are not driven by households becoming more risk-averse in general. For example, I find no evidence of spillover effects across other insurance products, such as earthquake or health insurance. Furthermore, in my second quasi-experiment (detailed below), I consider a random shock unrelated to risk aversion. Yet, I find a consistent effect of social interactions on households' insurance decisions. I also show that my findings cannot be explained by a story of consumption externalities, such as a desire to "keep up with the Joneses."

My second quasi-experiment follows the same stacked difference-in-differences design as above, but with a campaign (instead of a flooding event) as a random shock to geographically distant friends. The campaign is staggered across US counties, aiming to increase public awareness of flood risk. It has two components: first, it transforms an existing flood-risk map from black-and-white to colored;[9] second, when the new map is published in a county, the county government publicizes it in various ways (such as open houses and newspaper advertisements).

I first present evidence that the campaign is unexpected by the targeted households (in my setting, far-away friends). I then show that the campaign does affect their behaviors: the number of flood insurance policies in-force increases by 30.6%.[10] In addition, I document heterogeneous effects of the campaign; interestingly, counties with high intra-county social connectedness experience a stronger post-campaign surge in flood insurance purchases. This relationship is consistent with the effect of social interactions: in a more intra-connected county, households are probably more likely to have recursive discussions about flood risk—initially induced by the campaign—which could amplify its impact (Bali et al., 2018; Hirshleifer, 2020).

The core finding, however, is that insurance demand increases by 1.01% in counties that are more socially connected to the campaign county compared to the less-connected counterfactuals in the same distant state. Publicizing flood risk information that pertains only to far-away peers, the far-away campaign is orthogonal to local households' insurance

[9] Some areas of some maps might also have been updated, while being colorized, to reflect higher or lower flood risk, but the scope of the updating is small.

[10] There could be multiple channels through which the localized campaign affects the targeted households. For example, it could reduce information acquisition costs, provide financial education, or increase the salience of flood risk by attracting attention. This paper does not intend to differentiate them.

decisions. Therefore, this finding can only be attributed to the effect of social interactions. As in my first quasi-experiment, the effect is persistent, suggesting social learning as a key mechanism. To further validate the causal interpretation, I show that the treated and control counties—those more or less socially connected to the campaign county—do not have differential pre-trends in insurance purchases. I also find no evidence of correlated shocks; the probabilities of having a contemporaneous campaign in the treated and control counties are statistically identical and both close to zero.

This paper does not claim that the increase in flood insurance purchases—or that the NFIP in general—is socially optimal. It is likely to be 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.[11] A back-of-the-envelope calculation suggests that the expected net benefit of buying NFIP flood insurance is $1,240 per year per flood-prone household.[12] However, a complete analysis on social welfare must take government expenditures into account, which is beyond the scope of this paper.[13]

This paper contributes to the emerging field of social finance (Hirshleifer, 2020) which studies how social interactions shape economic and financial decision-making. Kuchler and Stroebel (2020) provide a review of recent empirical work at the intersection of social finance and household finance.[14] In particular, Bailey et al. (2018a) show that friends' house-price experiences affect one's own housing investment decisions. Several recent papers find that social networks play a prominent role in information dissemination in various markets. For example, Allen et al. (2020) show that social interactions facilitate the spread of new financial products (online lending marketplaces). Maturana and Nickerson (2019) and McCartney and Shah (2019) show that peer interactions affect individuals' choices of mortgage lenders and refinancing decisions, most likely through the channel of transmitting otherwise costly information. To the best of my knowledge,

[11]The NFIP suggests that eliminating the subsidy would cause aggregate premiums to increase by 50-75% (Hayes and Neal, 2011). Consistently, as premiums are set below actuarial levels, they cannot fully cover claims: the cumulative difference is –$5.85 billion over the past 20 years. The NFIP's 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).

[12]The estimation is based on the following assumptions: an NFIP-defined 1% inundation probability p.a., an average premium of $993, an average coverage of $227,112, and an average deductible of $3,831.

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

[14]Prior research on peer effects on household financial decisions includes: Duflo and Saez (2002, 2003); Grinblatt et al. (2008); Bursztyn et al. (2014); Beshears et al. (2015); Hvide and Östberg (2015). More broadly, for peer effects on financially affluent individuals, see: Cohen et al. (2008, 2010); Shue (2013); Pool et al. (2015); Kuchler et al. (2020).

my paper provides the first empirical evidence that social interactions influence insurance purchase against natural disasters, which is an increasingly critical financial decision faced by households. Thanks to the special setting of the US flood insurance market, my paper also differentiates from the previous studies by proposing and evaluating a channel of attention-triggered indirect social learning.

This paper also adds to the nascent but rapidly growing 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 (such as a day of unusual weather). The persistent effects of social interactions, documented in this paper, have unique implications for effective climate policy, given that offering subsidized insurance is costly yet has only limited success.

This paper is among the first to study households' insurance decisions against rare disaster risks. A closely related paper is Gallagher (2014). My work differs in several dimensions. First, Gallagher shows that local flood insurance purchases increase after a local flood, whereas my paper focuses on the effects of social interactions. Second, he documents a long-term but slowly diminishing effect of personal experiences, while I provide evidence of a persistent effect of social interactions, suggesting an important role of social learning.[15] Third, his NFIP data is yearly, aggregated, and covers an earlier period (1980–2007). Fourth, from the perspective of policy interventions, governments cannot create floods for households to experience, whereas my findings on social learning (and the direct campaign effect) offer clearer policy implications.

The remainder of the paper is organized as follows. Section 2 describes the insti-

[15] In Figure 2 of Gallagher (2014), the effect appears to be quite persistent (around 8%) in the first eight years after a flood, consistent with my finding that the renewal rate is 88.7% by the end of the eighth year in my setting. The effect in his setting then drops by half in year 9 and becomes statistically insignificant in the next two years. One possible explanation for the reversal is, in my view, that multiple factors might simultaneously change during a flood, some of which might be relatively short-term (e.g., a behavioral bias of overweighting recent experiences). Unfortunately, as my sample spans slightly less than ten years, I cannot evaluate the longer-term persistence of my results.

tutional background and details the data. Section 3 describes the empirical approach for identifying a causal effect of social interactions on households' flood insurance purchases. Sections 4 and 5 present findings from the use of two different random shocks to geographically distant friends. 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; as of 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); 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 house characteristics and choices of coverage. 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 and termination dates, premiums, coverage, deductibles, first policy dates, cancellation dates, and house characteristics. The policies are renewed annually; renewals appear as separate transactions in the dataset. Broad location information (such as census tract, county code, and community code) is available, but specific properties cannot be identified due to privacy protection (geographic coordinates are truncated to one decimal point).

From the policy effective 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 equals 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 have transaction data for 2008, I can not calculate 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 both zoom in on households' flood insurance demand in the very short-run and 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 premiums for $1.26 trillion in coverage.[16] The nationwide average annual premium per policy is $628, and the average coverage per policy is $238,000. Panel B shows that the cross-sectional heterogeneity is stark. While the average county purchases 1,766 policies, the median is only 120. Figure 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 Social Connectedness

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 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 in those two counties. They further create a measure called the *relative probability of friendship* by dividing the SCI for counties 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.

Following several other papers that use and validate the SCI (Bailey et al., 2018a; Bali et al., 2018; Bailey et al., 2019a,b; Wilson, 2019; Allen et al., 2020; Bailey et al., 2020; Kuchler et al., 2020), I argue that the SCI is a sensible proxy for real-world US social networks and for both online and offline social interactions. This is a result of Facebook's enormous scale (234 million active users in the US and Canada), its comprehensive market penetration, and the fact that people primarily use Facebook to interact with their real-world friends (whereas links to non-acquaintances are more common on other social media

[16]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).

platforms, such as Twitter and LinkedIn). Survey evidence provided by the Pew Research Center shows that 79% of online Americans use Facebook in 2016 (that is 68% of all US adults, accounting for Americans who do not use the Internet at all).[17]

2.3 Shock One: Flooding Events

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 start 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.

2.4 Shock Two: Flood-risk-map Campaigns

The Risk Mapping, Assessment, and Planning program—known as the Risk MAP—was launched in 2009 by FEMA, aiming to increase public awareness of flood risk. This campaign transforms existing black-and-white flood-risk maps into colored ones and publicizes them. Before the reform, 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.

2.4.1 Map Transformation

Figure 2 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-risk maps (available at msc.fema.gov/portal). Figure 2.a shows a scanned copy of the legacy paper map. It was published on November 16, 1995 and was in use until the new colored map (Figure 2.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 key information is the risk zone designation. The blue areas in Figure 2.b are the Special Flood Hazard

[17] See https://www.pewresearch.org/internet/2016/11/11/social-media-update-2016

Areas, which are expected to have a 1-percent 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-percent flooding probability per year (referred to as 500-year floodplains).

2.4.2 Local Campaigns

Around the publication of the new colorized 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.[18] Appendix Figure A.1 presents examples of actual advertisements and announcements.

2.4.3 Staggered Campaign Rollout

Revisiting Figure 2.b, the dates shown on each small area indicate the publication dates of the new flood risk maps. For instance, the Town of Colfax (along with other communities in the same county) published its new colored map on June 16, 2016, whereas the neighboring county to the west did so on July 6, 2015. The colored map for the dot-shaded area to the south-west (part of a different county) is not yet available; the latest version is still the scanned black-and-white 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 rollout happens at the county level; communities in 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 social connectedness, is not as granular as communities, I examine the staggered campaign at the county level. When there are disparities among communities, I define the campaign time of a county as the calendar month in which more than 50 percent of its communities simultaneously publish their new maps.[19] Appendix Figure A.2 maps the staggered rollout by time and county. The darker the shade, the more recent the campaign dates; unshaded areas did not have the

[18] 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.

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

campaign in the 2010s.[20]

3 Empirical Strategies

Peer effects are difficult to identify apart from the effects of selection in a friendship group and the exposure to common shocks, highlighted in Manski (1993). In this section, I discuss the empirical approach I use to measure the effect of peers on the decision to purchase flood insurance. My identification relies on exogenous shocks to an individual's far-away friends, which should not affect her own insurance decision in any other way than through peer effects from friends who receive the shock. To further address the concern of common shocks, I use a difference-in-differences strategy to compare households within the same state that are geographically distant from the shock.

3.1 Social Interactions and Geographically Distant Floods

I first consider non-local flooding events as the random shock to geographically distant friends, which is orthogonal to an individual's local flood risk pertaining to her own insurance decision. The change in the number of flood insurance policies in-force at the county level thus reflects the aggregation of the peer effects across individuals. More concretely, for a given flood f, I identify a set of flooded counties $\{j\}_f$ and a set of geographically distant counties $\{i\}_f$ (at least 750 miles away). 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 2.2):

$$p_{i,f} = \sum_{\{j\}_f} w_j * p_{i,j}, \tag{1}$$

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) the state median.

[20]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. However, Appendix Section A presents evidence suggesting that FEMA largely falls short of that goal. Hence, the date I observe is likely to be the county's only map publication in the 2010s.

Figure 3 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.

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

$$Y_{it} = \beta_0 + \beta_1 * Connected_i \times PostFlood_t + \beta_2 * Connected_i + \beta_3 * PostFlood_t + \epsilon_{it}. \quad (2)$$

$Connected_i$ is the treatment dummy. $PostFlood_t$ is the post-event dummy (that turns on if t is after the flooding time). The outcome variable 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. The interaction $Connected_i \times PostFlood_t$ is the key explanatory variable of interest.

By construction, I define treated and control counties conditional 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 (2) 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 must 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 $PostFlood_t$ with a sequence of event time dummies $\{\mathbb{1}(t = t^* + k)\}$ (t^* denotes the flood time). 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 4.1. I discuss potential concerns about this event study approach in Section 4.5 and show that my findings are robust to an alternative empirical methodology.

3.2 Social Interactions and Geographically Distant Campaigns

Following the same empirical framework as above, my second quasi-experiment leverages the NFIP's flood-risk-map campaigns (described in Section 2.4) as the shock to geographically distant friends. Unlike the previous setting, for a given event c, the set of campaign counties $\{j\}_c$ only contains a single county j. The rest is similar: I identify a set of counties $\{i\}_c$, which are at least 750 miles away from the campaign county j; within a state, county i is coded as treated (control) with respect to event c, if $p_{i,j}$ is above (below) its state median; the coefficient β_1 of the interaction $Connected_i \times PostCampaign_t$, in the following Regression (3), captures the effect of peers:

$$Y_{it} = \beta_0 + \beta_1 * Connected_i \times PostCampaign_t + \beta_2 * Connected_i + \beta_3 * PostCampaign_t + \epsilon_{it}. \quad (3)$$

$Connected_i$ is the treatment dummy, which equals 1 if county i has above-state-median connectedness to the campaign county j. $PostCampaign_t$ is the post-event dummy, which equals 1 if t is after the campaign time. The outcome variable Y_{it} is the normalized number of flood insurance policies in-force, as defined in Regression (2). I stack individual event studies (with respect to every campaign event) and cluster the standard errors at the county level.

4 Empirical Findings: Geographically Distant Floods

In the following two sections, I test the hypothesis that households purchase flood insurance because of social interactions. Section 4 presents the results of my first quasi-experiment, which exploits non-local flooding events that occur to geographically distant friends. I also consider several alternative explanations for my findings and present theoretical and empirical arguments to characterize their relevance.

4.1 Main Results

Using the event study framework detailed in Section 3.1, I compare the changes in flood insurance policies in-force across counties in the same state with high versus low social connectedness to a geographically distant flooded area. The connected (i.e., treated) counties are defined as having a connectedness measure above the state median.

Table 2 reports the coefficient estimates of Regression (2). To reject the null hypothesis that social interactions are irrelevant, we expect a positive difference-in-differences estimate. In column 1 of Panel A, the estimate of 1.11 represents a 0.94-percent increase over the average number of policies in-force (117.66) at event time zero. It suggests that more households decide to acquire flood insurance when they have geographically distant friends who experience a flood.

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

In Panel B of Table 2, I further check the robustness of my results. Since the Facebook data (described in Section 2.2) 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 social dynamics back in the early 2010s. Panel B of Table 2 therefore examines a period between 2014 and 2017 to address this concern. It shows that my findings are not sensitive to the choice of sample period.

Figure 4 shows the dynamic effects. Crucially, there is no evidence of a differential pre-trend. This supports the identifying assumption that insurance purchases in the more- and less-connected counties (with respect to a far-away flood) would have evolved in parallel. Figure 4 also suggests that the effect is persistent; the number of policies in-force remains stable and does not revert. Recall that the policies are renewed annually; thus, comparing event months 1–6 with event months 13–18, for example, we could posit that the additional purchases induced by the shock are renewed one year later. Summary statistics of renewals are consistent with this assertion: for those first-time new policies purchased within a year of a distant flood, the average "survival" rates are 97.3%, 93.1%, 90.4%, and 88.7%, after 2, 4, 6, and 8 years, respectively.

4.2 Heterogeneous Effects

4.2.1 Monotonicity in Social Connectedness

My hypothesis implies that the effect of social interactions should be monotonic in the strength of social connectedness. The most-connected (least-connected) counties should show the largest (smallest) increases in insurance demand.

Recall that my baseline analysis (in Table 2) defines the treatment or control group as respectively above or below the state-median value of the social connectedness measure, as per Equation (1). In Panel A of Table 3, I use a sharper approach and compare the top versus bottom quartiles. The estimates are about twice as large in magnitude as the baseline results, 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 impact of friends' flood experiences, transmitted through social interactions via social networks, is indeed monotonic in the strength of social connectedness.

4.2.2 Significant Floods

My hypothesis also implies that the most damaging floods, which create the strongest shocks to peers, should cause the most pronounced effect across social networks. Panel B of Table 3 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 two times larger than the baseline in Table 2. 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 percent) in my analysis.

4.3 Non-causal Explanations and Further Evidence

In order to interpret the previous results as the causal effect of social interactions, I rule out potential non-causal mechanisms that might also induce the observed correlation between households' flood insurance decisions and their friends' flood experiences.

4.3.1 Common Shock and Homophily

Endogenously formed peer groups are subject to common shocks and homophily—people are more likely to be friends with others who have similar characteristics and preferences (McPherson et al., 2001). As a result, correlated behavior across friends does not necessarily imply the presence of peer effects.

For example, two neighbors who are friends might simultaneously experience a flood or a campaign for insurance take-up, which could induce both to acquire flood insurance quite independently. An example of homophily could be that two individuals who both

[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.

like living near the water and thus have flood insurance independently may be more likely to become friends (e.g., through participating in aquatic sports together).

My identification strategy specifically attempts to address these concerns in several ways. First, a flood occurrence is quasi-random by nature. Second, I focus on floods experienced by geographically distant friends, which are unlikely to systematically coincide with a local flood. Third, the difference-in-differences construct of my empirical design further minimizes the possibility of correlated flood shocks. To violate the identifying assumption, the alternative mechanism requires that the more-connected-to-Boston California counties, for example, contemporaneously experience a flood, while the less-connected-to-Boston California counties do not. This is unlikely, both intuitively and consistent with the evidence presented below. Fourth, my identification does not require the absence of homophily. For example, it is not necessarily a problem that two counties with high flood insurance coverage, despite being far apart, are more socially connected for some reason. The difference-in-differences estimate teases out the fixed differences, such as the inherent flood risk, between the more- and less-connected counties.

Furthermore, I provide evidence to cast doubt on this alternative explanation. First, Figure 5 shows that there is no difference in flooding frequency between the more- and less-connected counties, before and after the distant shock. If the alternative hypothesis were true (i.e., if there were correlated flood shocks), we would expect to see bunching of floods around event time zero; we do not. Second, I present evidence that the more- and less-connected counties have similar inherent flood risk. To measure risk level, I leverage the Special Flood Hazard Area (SFHA) defined by FEMA, which is expected to have a one-percent or higher probability of being inundated in any given year. I calculate the proportion of SFHA policies in a county as a proxy for its average underlying flood risk and find that the more- and less-connected counties have similar SFHA-proportions—53.8 percent versus 55.5 percent. Third, in unreported results (for brevity), I also find that they are similar in terms of average age, household income, and education (using US Census Bureau data).

4.3.2 Migration

A second concern is that households may move after a flood, plausibly to places where families and friends live, and subsequently purchase flood insurance for their new homes without necessarily influencing their local peers. This alternative explanation predicts that the more-connected counties receive more incoming households migrating from the flooded area than the less-connected counties. To assess its empirical relevance, I use the county-to-county migration data produced by the Internal Revenue Service (IRS), which is based on year-to-year address changes reported on individual income tax returns filed with the IRS.

The data suggest that long-distance migration is uncommon and hence unlikely to be the driving factor of my results. The average number of households migrating from a flooded area to a 750-mile-away county, within one year of the flood, is only 3.6 (the median is 0). The magnitude is small compared to the average number of households (35,637) in a county. Moreover, there is little difference between the more- and less-connected counties. Using the number of migrating households as the dependent variable in Regression (2), the difference-in-differences estimate is 0.99 (t-statistic=0.94). The estimate is economically and statistically insignificant, suggesting that the increase in insurance demand in geographically distant counties is not driven by migration.

4.4 Social Interactions and Insurance Decisions: Mechanisms

In the previous section, I document a causal relationship between social interactions (with geographically distant friends experiencing floods) and households' flood insurance decisions. Here, I investigate potential mechanisms behind this causal relationship.

4.4.1 Proposed Mechanism: Social Learning

Households may update their beliefs about flood risk and flood insurance via social interactions with peers. There are two potential forms of social learning—direct and indirect learning—and I provide theoretical and empirical arguments to weigh their relevance in my specific quasi-experiment. While both learning mechanisms imply that households acquire new information as a consequence of interacting with peers, it is useful to conceptually distinguish them, even though empirical tests are somewhat limited by the data.

Direct Channel

Social interactions may act as an informational source that provides pertinent and otherwise costly information. The direct channel could be particularly strong in markets with high information asymmetries and when other possible sources of information (such as mortgage brokers) may have conflicts of interest. For example, Maturana and Nickerson (2019) suggest that teachers are likely to acquire financial information from fellow teachers about the benefits of mortgage refinancing, the prevailing rate, and the lenders offering these terms.

For flood insurance, households may acquire or infer information from peers about either the flood risk of their local neighborhoods or the insurance products offered by the NFIP. This information is pertinent to one's insurance purchase decision.

In my specific empirical design, direct information dissemination is probably mild and I provide some suggestive arguments. First, as my analysis leverages on geographically distant friends' flood experiences, households are unlikely to directly infer information about their local flood risk during such events, as the occurrences of far-away floods are orthogonal to local flood risk (Figure 5 supports this argument).

Second, the US flood insurance market is subject to arguably minor information frictions, as it is dominated by the NFIP which has offered standardized insurance policies with fixed rates for over 50 years. As described in Section 2, households can easily obtain information about the NFIP and about their own properties' flood risk in multiple ways at almost no cost. Therefore, households are less likely to rely on information from friends than they would in other markets (e.g., choosing mortgage brokers).

Third, in Appendix Section B, I provide survey evidence that when experiencing a flood, individuals are much more likely to make Facebook posts about the flood itself (such as pictures and videos) than about flood insurance. However, I point out the caveat that this analysis is limited to only interactions on Facebook and is thus unable to preclude the possibilities that friends have offline discussions about the importance of flood insurance. It is important to note that the networks formed on Facebook should best be viewed as a proxy for broader social interactions: connected individuals are likely to interact in general whether on or off the platform.

Indirect Channel

Households being subject to limited cognitive resources and attention, social interactions may nudge attention to freely available but less-salient public information. In my context, even if the interactions with far-away friends experiencing a flood may not transmit direct information on local flood risk or flood insurance, they can still facilitate social learning indirectly. Interacting with friends having flood experiences could turn an individual's attention to flood risk and thus enable her to gather information about her own risk that had always been freely available in the public information set but to which she had previously been inattentive. The persistent effect of social interactions (Figure 4) is consistent with a learning process which leads to a permanent shift in insurance take-up.

4.4.2 Alternative Mechanism: Risk Aversion

A first alternative causal mechanism of social interactions influencing households' insurance decisions is by affecting their risk attitude. It could be that individuals become generally more risk-averse after seeing pictures of natural disasters shared by friends. If so, we would expect to see households purchase more insurance in general to hedge various types of risks. However, I find no spillover effect across other insurance products, suggesting that this mechanism cannot explain my findings.

First, I examine the effect of geographically distant friends' flood experiences on households' purchases of health insurance. I obtain county-year-level panel data from the US Census Bureau's Small Area Health Insurance Estimates (SAHIE) program from 2010 to 2018. I use health insurance coverage (in percent) from the SAHIE dataset as the dependent variable in Regression (2), with the time dimension collapsed to yearly. Panel A of Table 4 reports the results, which suggest that geographically distant friends' flood experiences do not induce the purchase of health insurance. For example, in column 1, the coefficient estimate of $Connected_i \times PostFlood_t$ suggests that the counties more connected to a distant flooded area experience a 0.012-percentage-point decrease in health insurance coverage relative to the less-connected counties in the same state. The effect is statistically and economically insignificant (the unconditional mean of health insurance coverage is 84.9 percent).

Second, in Panel B of Table 4, I investigate changes in earthquake insurance take-up after a geographically distant flooding event. Due to data limitations, I am only able to obtain county-year-level information on earthquake insurance purchases for one state—Missouri—from its Department of Insurance. I use the normalized number of earthquake insurance policies in-force as the dependent variable in Regression (2). The results presented in Panel B suggest that distant friends' flood experiences have little impact on households' purchases of earthquake insurance.

Third, I consider a placebo test to further evaluate a potential spillover effect across insurance products. Specifically, I keep the dependent variable in Regression (2) as the number of flood insurance policies in-force, but construct the event study using non-flood-related natural disasters. According to the Presidential Disaster Declaration database (see Section 2.3), there were 222 non-flood-related incidents (such as wildfires, earthquakes, and tornadoes) during my sample period. Panel C of Table 4 presents the effect of geographically distant friends' experiences of non-flood disasters on households' purchases of flood insurance. I find no significant spillover effect, which is inconsistent with the risk-aversion channel.

4.4.3 Alternative Mechanism: Consumption Externalities

A second potential alternative causal explanation for my findings is the presence of consumption externalities in insurance decisions. Gallagher (2014) shows that flood insurance purchases increase in the flooded areas; it may therefore be that an individual, living far away from the flood, derives utility from mimicking her distant friends' behaviors. If the desire to "keep up with the Joneses" were the only motive for buying flood insurance, we would expect my results to be homogeneous across geographies with different underlying flood risk.

To test this hypothesis, I restrict the "experiment" states, according to the inherent risk levels, into quintiles and estimate Regression (2) for each subsample. Figure 6 plots the coefficient estimates of the effects of far-away friends' flood experiences on the number of flood insurance policies in-force, in states ranked from low-risk to high-risk. The effects appear to be non-linear—in particular, inverse U-shaped. This result cannot be explained purely by the desire to keep up with friends; instead, it is more consistent with the mechanism of social learning. The fact that the strength of the effect depends (roughly positively) on the underlying risk suggests that learning plays an important role and that risk information is a crucial factor in households' insurance decisions. In addition, the mechanism of attention-triggered indirect social learning might explain why the effect is not as strong in the top quintile: households in high-risk areas may already be fully attentive to flood insurance.

4.5 Alternative Methodology

In this section, I address a concern of my event-study design to link social connectedness with flood insurance purchases. I show that my findings are robust to an alternative empirical methodology.

The advantage of my strategy is that each flood characterizes a standard difference-in-differences analysis, allowing for 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 event, which entails duplicating observations if the event windows overlap. In other words, instead of one observation per county per time, I have 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 all relevant tables). 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 3, with only the largest 18 floods included, is undertaken in this spirit.

Furthermore, I use an alternative empirical approach proposed 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 might 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:

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

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 in 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:

$$log(Policies)_{i,t} = \beta * FriendFlood^{Distant}_{i,t-12,t} + FE_{state\times time} + \epsilon_{i,t}. \quad (4)$$

By controlling for the state×time fixed effects, I can isolate the effects of friends' flood experiences on the insurance decisions made in counties in the same state at the same time.

Panel A of Table 5 presents results from Regression (4). 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 for floods. Columns 5 through 7 show that my result is also insensitive to a variety of definitions of "geographically distant." These results are similar to my main findings presented in Table 2.

Panel B of Table 5 repeats the analysis by focusing on significant floods only (as previously defined in Section 4.4.2). Across all columns, the estimates in Panel B are more than twice as large as those 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%.

5 Empirical Findings: Flood-risk-map Campaign

To further validate the effect of social interactions on insurance decisions, I exploit an alternative shock to geographically distant friends. The shock is characterized by a staggered localized flood-risk-map campaign across US counties, as detailed in Section 2.4, aiming to increase public awareness of flood risk. My findings support the hypothesis that, given a far-away campaign orthogonal to local risk, flood insurance demand nonetheless increases because of social interactions.

5.1 An Unexpected Shock to Friends

I begin this section by documenting that the flood-risk-map campaign does affect households' behaviors; specifically, they purchase more flood insurance. Next, I present evidence that this campaign is, by and large, unexpected. These two pieces of evidence validate my empirical design and the main findings presented in Section 5.2.

As previously illustrated in Appendix Figure A.2, the campaign for new colorized flood-risk maps is staggered across US counties. Leveraging variation in the campaign timings, I estimate the following canonical staggered difference-in-differences regression:

$$Y_{it} = \alpha_i + \lambda_t + \beta * Campaign_{it} + \epsilon_{it}. \tag{5}$$

In this two-way fixed-effects model, unit and time are specified as county and year-month. Y_{it} measures, as in Regression (2), the normalized number of policies in-force in county i at time t. Let t_i^* denote the campaign time for county i. $Campaign_{it}$ is an indicator variable $\mathbb{1}(t > t_i^*)$ that turns on if county i has had the campaign at time t; it 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 treatment.[22]

Table 6 reports the results of Regression (5). Column 1 shows that following the campaign, the average number of insurance policies goes up dramatically by 21.39 in the targeted county. Note that this is the mean effect over the entire post period; it is 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 average premium per policy, average coverage per policy, and number of flood incidents. In column 3, I exclude the never-treated counties from my analysis; thus, to control for any underlying trends, the staggered difference-in-differences estimation uses only counties that have not yet had or have already had the campaign. In both columns, the coefficient estimate has a similar magnitude and statistical significance.

In column 4, I present the heterogeneous effects related to social interactions. To assess if households with different intra-county connectedness react differently to the campaign, I add an interaction term—between the post-treatment indicator $Campaign_{it}$ and $HighSCI_i$—into the regression. $HighSCI_i$ is a binary variable that equals 1 if county i has an above-median value of the intra-county *relative probability of friendship* (that is,

[22]The recent literature in applied econometrics shows that two-way fixed-effects estimations of difference-in-differences coefficients can lead to substantial biases when there are staggered treatment timing and heterogeneous/time-varying treatment effects (Baker et al., 2021; Borusyak et al., 2021; Callaway and Sant'Anna, 2021; De Chaisemartin and d'Haultfoeuille, 2020; Goodman-Bacon, 2021; Sun and Abraham, 2021). I follow the suggestions from the literature and show that the bias is small in my setting. See Appendix Section C for details.

the $p_{i,i}$ described in Section 2.2). The positive coefficient of the interaction term suggests that the more intra-connected counties are more responsive to the campaign, potentially because households in these counties are more likely to interact and to have recursive discussions about flood risk, leading to amplification of the campaign's impact.

The identifying assumption of Regression (5) requires the treatment (i.e., campaign) timing to be uncorrelated with the outcome (i.e., insurance purchases). If this assumption is not satisfied, the treated counties might already diverge from the controls before treatment. To assess the pattern of households' insurance decisions relative to the campaign date, I estimate the following nonparametric model:

$$Y_{it} = \alpha_i + \lambda_t + \sum_{k=-\underline{L}}^{k=\overline{L}} \beta_k * \mathbb{1}(t = t_i^* + k) + \epsilon_{it}, \tag{6}$$

where $\{\beta_k\}$ for $k < 0$ correspond to the pre-trends and $\{\beta_k\}$ for $k > 0$ measure the campaign's dynamic effects, relative to the omitted β_{-1}.

Figure 7 depicts the results of Regression (6). Crucially, there is no evidence of a differential pre-trend in the campaign counties, suggesting that without the campaign, flood insurance purchases would have moved in parallel in the treated and untreated counties. Following the campaign, insurance purchases rise sharply: by the end of the first year, the number of policies in-force increases by 32.8 (a 30.6-percent increase over the mean of 107.4 at event time zero).

5.2 The Effect Transmitted via Social Networks

In this section, I examine the impact of the campaign on households in distant areas. Specifically, I take a given campaign as a shock and focus on counties in a far-away state. Within that state, I compare changes in flood insurance purchases between counties that are more and less socially connected to the campaign county.

Table 7 presents the estimates of Regression (3). To reject the null hypothesis of the irrelevance of social interactions, we expect a positive difference-in-differences estimate. Column 1 shows that following the flood-risk-map campaign in a geographically distant county, the average number of flood insurance policies in-force increases by 1.21 in the more-connected counties (relative to the less-connected), a 1.01-percent increase relative to the mean of 120.1 at event time zero. Columns 2 to 4 use various distance thresholds to construct my analysis; results are robust to these alternative specifications.

Figure 8 plots the dynamic effects. It is reassuring that the identifying assumption of parallel pre-trends is not violated, which supports a causal interpretation of the positive relationship between social interactions and flood insurance purchases. To further validate causality and to rule out the possibility of correlated shocks, I show that the treated and control counties are indeed similar in many observable dimensions and differ only in social connectedness to the given far-away campaign. For example, Appendix Figure A.3 suggests that the increase in insurance demand in the more-connected counties is not driven by contemporaneous local campaigns. In unreported results (for brevity), I also check that treated and control counties are similar in flood occurrence, underlying flood risk, and various demographics.

Figure 8 suggests that local households purchase roughly one percent more flood insurance policies, following their friends' experiences of the flood-risk-map campaign in a far-away location. The effect appears to be persistent, interestingly resembling the pattern in Figure 4 from my first quasi-experiment (although the two experiments do not necessarily need to produce similar results). The persistent pattern is consistent with the mechanism of social learning through which the insurance take-up rate reaches its new long-run steady state. In this context, while the campaign publicizes flood-risk information relevant only to a far-away location, indirect social learning can still be viable through social interactions. Hearing friends talking about flood risk and their campaign experiences is a natural stimulus to find out one's own exposure to flood risk, even if such interactions do not immediately provide that information.

Furthermore, the difference between Figure 7 (the intended effect of the campaign) and Figure 8 (the effect of social interactions) is evident. The response in the campaign county is not only stronger but also materializes more quickly; the largest spike in the number of new purchases occurs immediately after the campaign. In contrast, the transition in Figure 8 is slower, suggesting a longer period of learning through distant social networks than through local connections.

6 Conclusion

This paper examines how social interactions shape households' insurance decisions against flood risk. To measure insurance demand, I use new transaction-level data from the

National Flood Insurance Program, which covers almost the entire US flood insurance market. I use two quasi-random shocks to geographically distant friends to identify the causal effect of social interactions on households' demand for flood insurance. I find that households increase insurance purchases by one to five percent after their geographically distant friends experience either a flood or a flood-risk-map campaign. My results are consistent with the hypothesis that peer effects facilitate social learning. This insight could be widely generalized to other types of tail risk (especially natural disaster risk) and to other countries.

References

L. Allen, L. Peng, and Y. Shan. Social networks and credit allocation on fintech lending platforms. *Available at SSRN*, 2020.

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. Bailey, E. Dávila, T. Kuchler, and J. Stroebel. House price beliefs and mortgage leverage choice. *The Review of Economic Studies*, 86(6):2403–2452, 2019a.

M. Bailey, D. M. Johnston, T. Kuchler, J. Stroebel, and A. Wong. Peer effects in product adoption. *National Bureau of Economic Research*, 2019b.

M. Bailey, P. Farrell, T. Kuchler, and J. Stroebel. Social connectedness in urban areas. *Journal of Urban Economics*, 118:103264, 2020.

A. Baker, D. F. Larcker, and C. C. Wang. How much should we trust staggered difference-in-differences estimates? *Available at SSRN 3794018*, 2021.

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.

T. G. Bali, D. A. Hirshleifer, L. Peng, and Y. Tang. Attention, social interaction, and investor attraction to lottery stocks. In *9th Miami Behavioral Finance Conference*, 2018.

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.

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.

J. Beshears, J. J. Choi, D. Laibson, B. C. Madrian, and K. L. Milkman. The effect of providing peer information on retirement savings decisions. *The Journal of finance*, 70 (3):1161–1201, 2015.

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.

K. Borusyak, X. Jaravel, and J. Spiess. Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*, 2021.

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.

L. Bursztyn, F. Ederer, B. Ferman, and N. Yuchtman. Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions. *Econometrica*, 82(4):1273–1301, 2014.

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.

B. Callaway and P. H. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 2021.

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.

J. Chivers and N. E. Flores. Market failure in information: the national flood insurance program. *Land Economics*, 78(4):515–521, 2002.

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

L. Cohen, A. Frazzini, and C. Malloy. The small world of investing: Board connections and mutual fund returns. *Journal of Political Economy*, 116(5):951–979, 2008.

L. Cohen, A. Frazzini, and C. Malloy. Sell-side school ties. *The Journal of Finance*, 65 (4):1409–1437, 2010.

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.

C. De Chaisemartin and X. d'Haultfoeuille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96, 2020.

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.

E. Duflo and E. Saez. Participation and investment decisions in a retirement plan: The influence of colleagues' choices. *Journal of public Economics*, 85(1):121–148, 2002.

E. Duflo and E. Saez. The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *The Quarterly journal of economics*, 118(3):815–842, 2003.

J. B. Elsner and B. H. Bossak. Bayesian analysis of us hurricane climate. *Journal of Climate*, 14(23):4341–4350, 2001.

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.

A. Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 2021.

M. Grinblatt, M. Keloharju, and S. Ikäheimo. Social influence and consumption: Evidence from the automobile purchases of neighbors. *The review of Economics and Statistics*, 90(4):735–753, 2008.

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. Hirshleifer. Presidential address: Social transmission bias in economics and finance. *The Journal of Finance*, 75(4):1779–1831, 2020.

D. Hirshleifer and S. H. Teoh. Limited attention, information disclosure, and financial reporting. *Journal of accounting and economics*, 36(1-3):337–386, 2003.

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

H. K. Hvide and P. Östberg. Social interaction at work. *Journal of Financial Economics*, 117(3):628–652, 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.

C. Kousky, H. Kunreuther, B. Lingle, and L. Shabman. The emerging private residential flood insurance market in the united states. *Wharton Risk Management and Decision Processes Center*, 2018.

T. Kuchler and J. Stroebel. Social finance. *National Bureau of Economic Research*, 2020.

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

T. Kuchler, Y. Li, L. Peng, J. Stroebel, and D. Zhou. Social proximity to capital: Implications for investors and firms. *National Bureau of Economic Research*, 2020.

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.

C. W. Landsea, B. A. Harper, K. Hoarau, and J. A. Knaff. Can we detect trends in extreme tropical cyclones? *Science*, 313(5786):452–454, 2006.

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 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.

C. F. Manski. Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542, 1993.

G. Maturana and J. Nickerson. Teachers teaching teachers: The role of workplace peer effects in financial decisions. *The Review of Financial Studies*, 32(10):3920–3957, 2019.

W. B. McCartney and A. Shah. The economic importance of neighbors: Evidence from hyperlocal social influence effects in mortgage markets. *Available at SSRN 2882317*, 2019.

M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

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.

R. A. Pielke. Are there trends in hurricane destruction? *Nature*, 438(7071):E11–E11, 2005.

R. A. Pielke, C. Landsea, M. Mayfield, J. Layer, and R. Pasch. Hurricanes and global warming. *Bulletin of the American Meteorological Society*, 86(11):1571–1576, 2005.

V. K. Pool, N. Stoffman, and S. E. Yonker. The people in your neighborhood: Social interactions and mutual fund portfolios. *The Journal of Finance*, 70(6):2679–2732, 2015.

J. C. Pope. Do seller disclosures affect property values? buyer information and the hedonic model. *Land Economics*, 84(4):551–572, 2008.

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.

R. J. Shiller. Narrative economics. *American Economic Review*, 107(4):967–1004, 2017.

R. J. Shiller. *Narrative economics: How stories go viral and drive major economic events*. Princeton University Press, 2020.

K. Shue. Executive networks and firm policies: Evidence from the random assignment of mba peers. *The Review of Financial Studies*, 26(6):1401–1442, 2013.

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.

B. Stevenson and J. Wolfers. Bargaining in the shadow of the law: Divorce laws and family distress. *The Quarterly Journal of Economics*, 121(1):267–288, 2006.

L. Sun and S. Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199, 2021.

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.

R. Wilson. The impact of social networks on eitc claiming behavior. *The Review of Economics and Statistics*, pages 1–45, 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

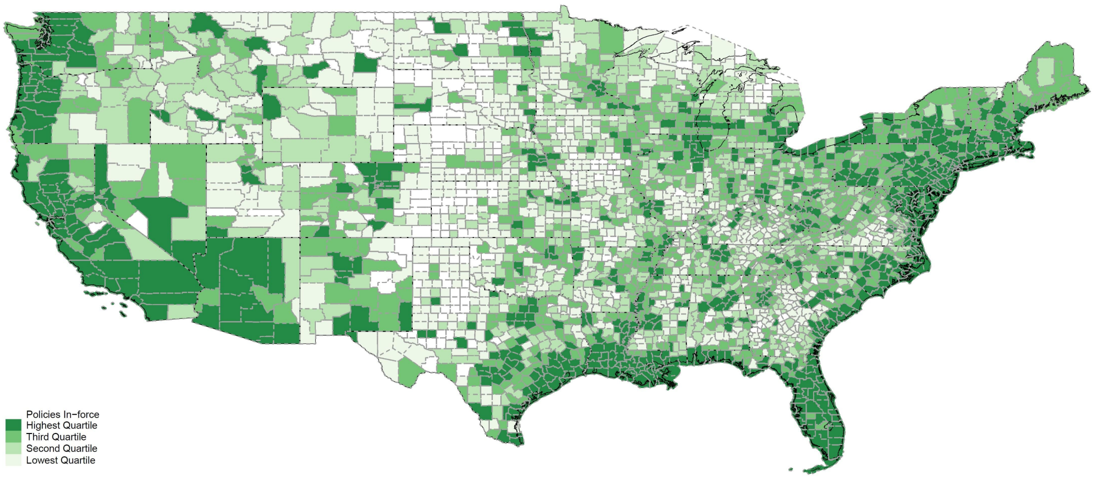
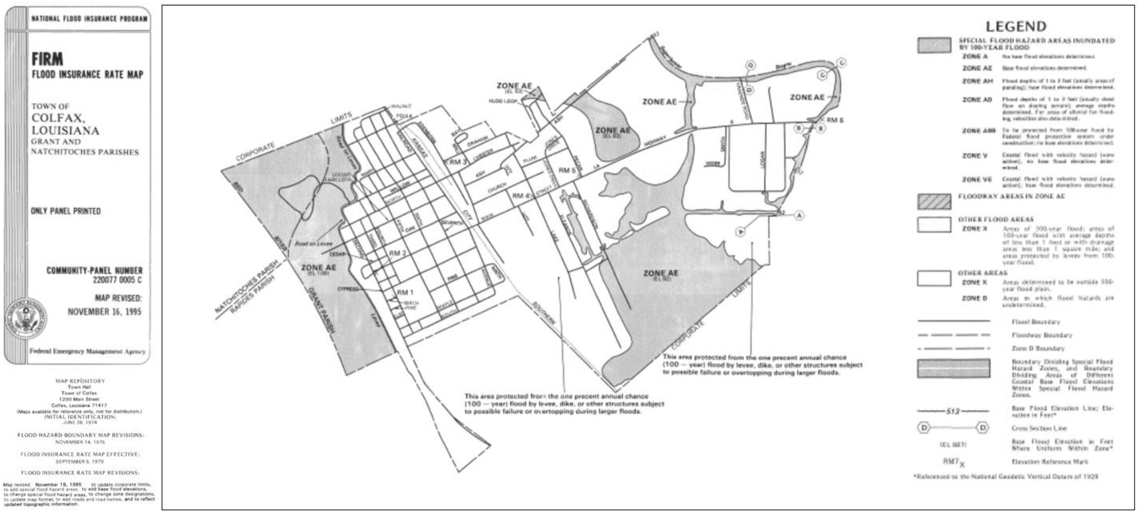
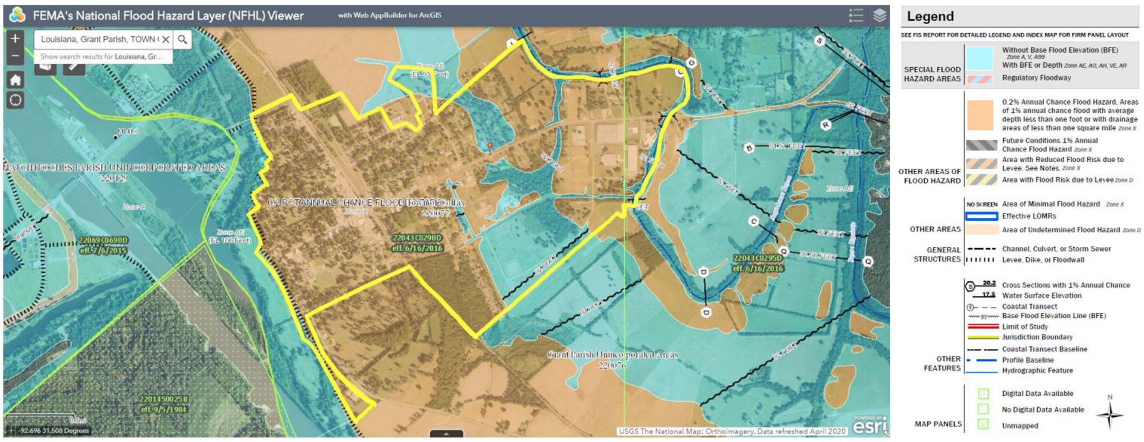


Figure 1. The Number of Policies In-force (Averages of 2010-2019)

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.



(a) Black-and-white Map Published on November 16, 1995



(b) Colorized Map Published on June 16, 2016

Figure 2. 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 black-and-white 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 colorized 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).

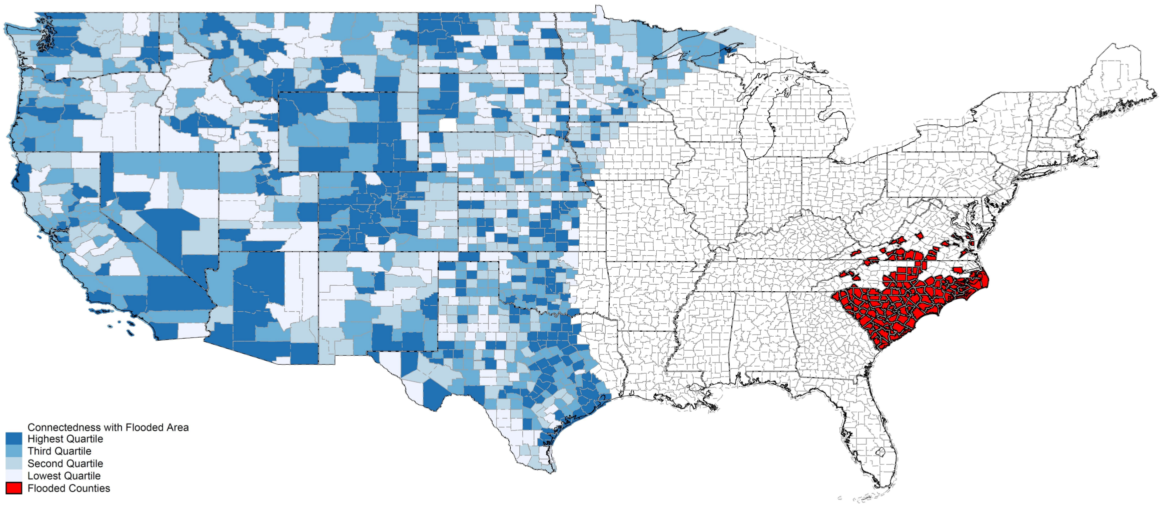


Figure 3. Social Connectedness and Geographically Distant Floods

This figure shows one specific example to illustrate the empirical design of my quasi-experiment. Hurricane Florence hit South Carolina, North Carolina, and Virginia in September 2018. The flooded 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 flooded area are considered.

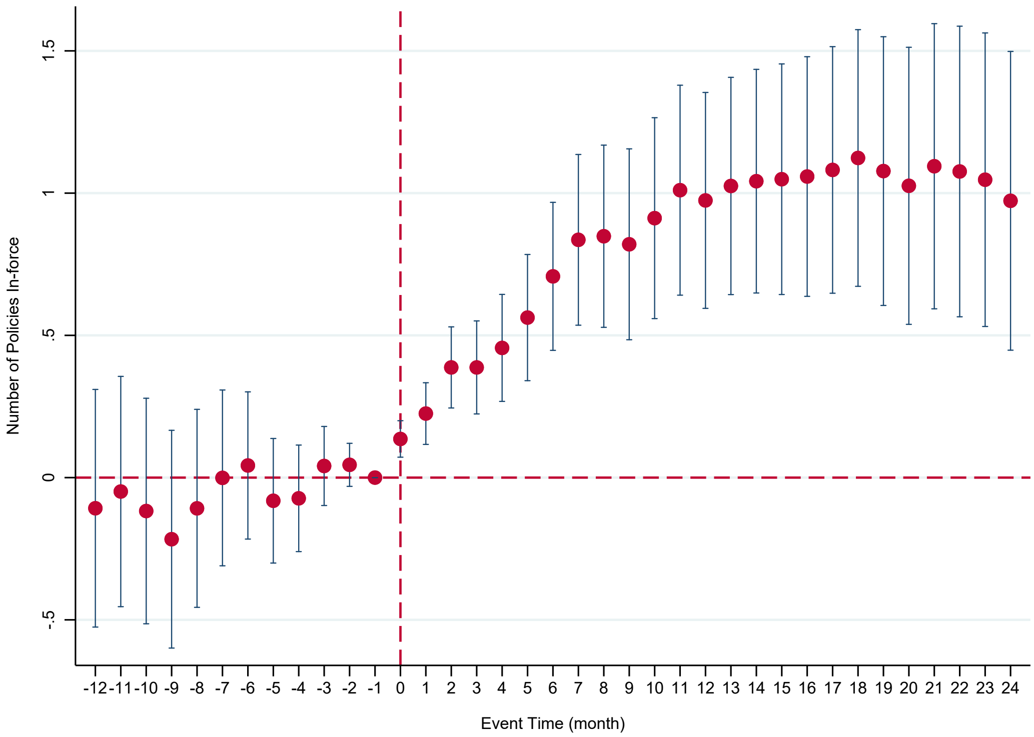


Figure 4. The Effect of Friends' Flood Experiences on Insurance Purchases

This figure shows the dynamic effects of geographically distant friends' flood experiences on insurance purchases. It plots the coefficient estimates of $\{\beta_1^k\}$ 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 f and the associated flooded counties $\{j\}_f$, $Connected_i$ is a binary variable indicating if county i has a above-state-median social connectedness with the flooded area. t^* is the month when the geographically distant flood occurs. The analysis sample consists of only counties that are at least 750 miles away from the flooded 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.

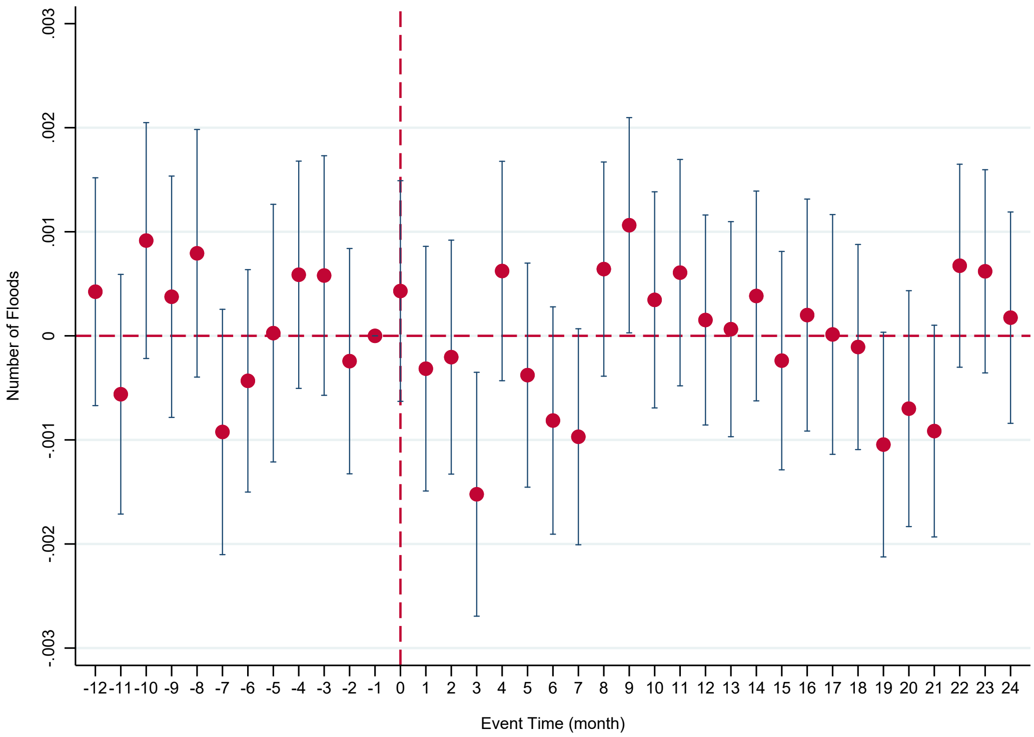


Figure 5. The Difference between Flood Occurrences in the Treated and Control Counties

This figure shows the dynamic effects of the geographically distant flood on local flood occurrences. It plots the coefficient estimates of $\{\beta_1^k\}$ in the regression as defined in Figure 4 except that the dependent variable Y_{it} measures the number of floods in county i in month t. Other variables are defined as per Figure 4. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.

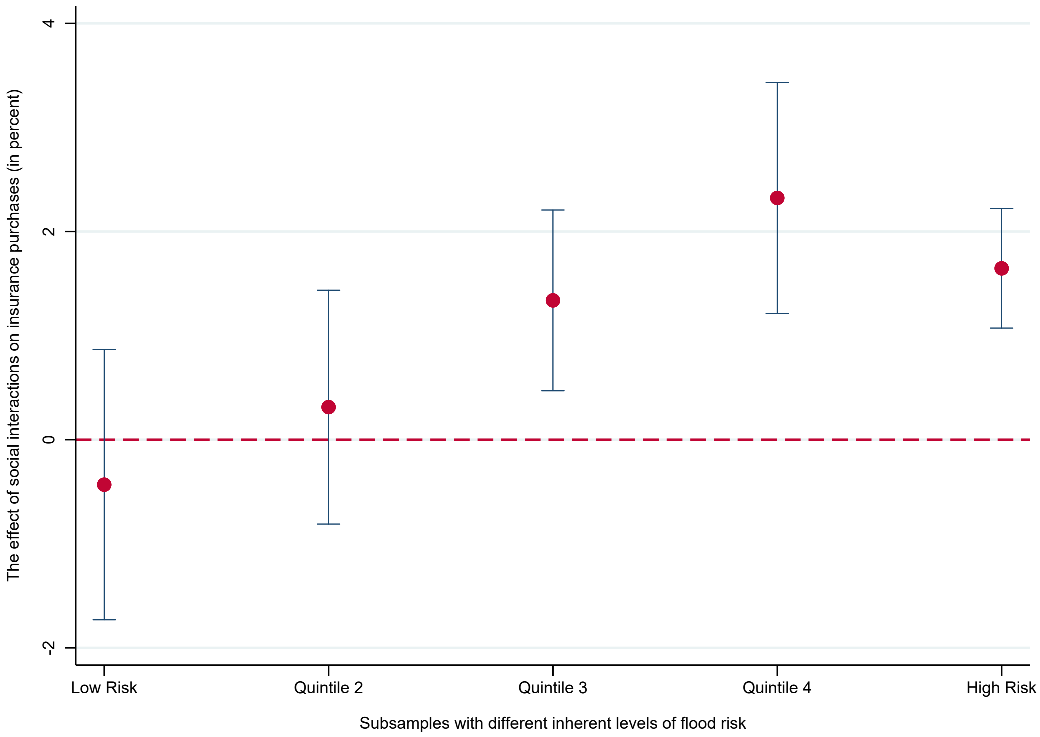


Figure 6. Heterogeneous Effects across Subsamples with Different Flood Risk

This figure shows the coefficient estimate of β_1 from the regression: $Y_{it}^{f} = \beta_0 + \beta_1 * Connected_i^f \times PostFlood_t^f + \beta_2 * Connected_i^f + \beta_3 * PostFlood_t^f + \epsilon_{it}^f$, using subsamples. Subsamples comprise states that are sorted into quintiles based on their average inherent flood risk. All variables are defined as in Table 2. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.

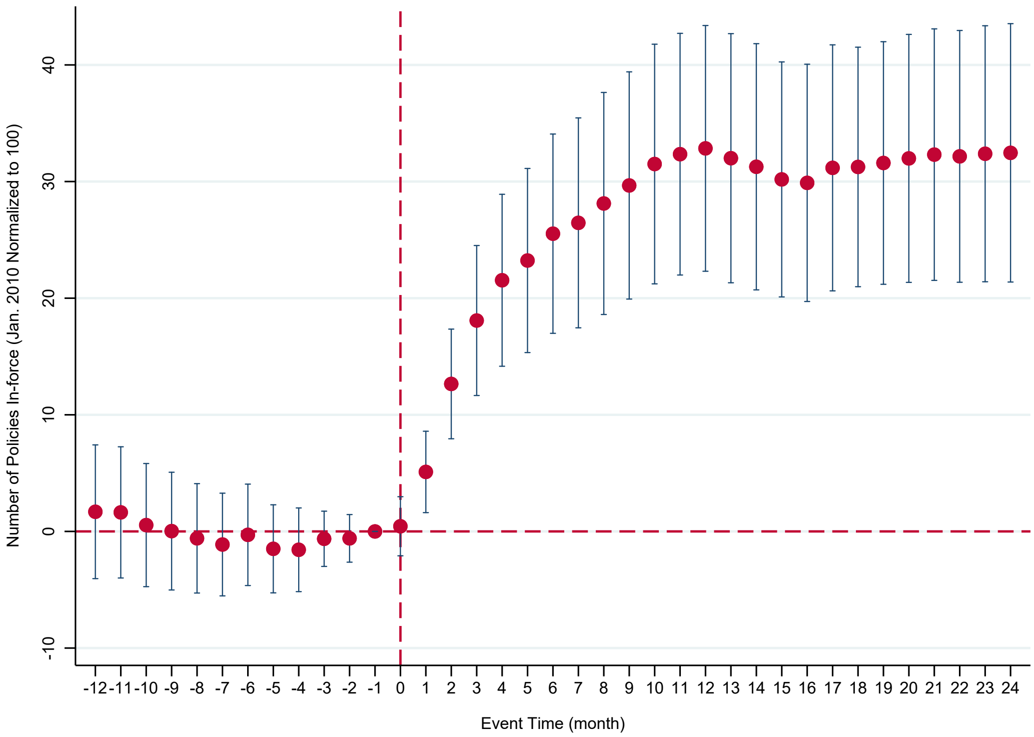


Figure 7. The Impact of Flood-risk-map Campaign on Insurance Policies In-force

This figure shows the dynamic effects of flood-risk-map campaign on insurance purchases in the targeted counties. 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 new 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.

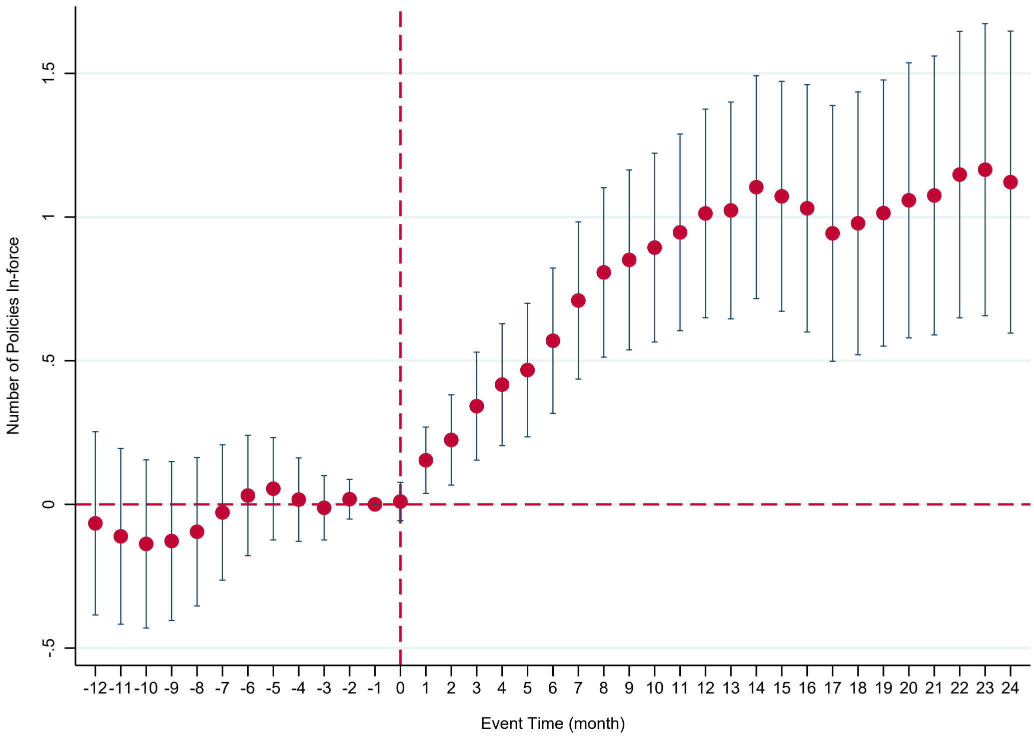


Figure 8. The Effect of Friends' Campaign Experiences on Insurance Purchases

This figure shows the dynamic effects of geographically distant friends' experiences of the flood-risk-map campaigns on insurance purchases. It plots the coefficient estimates of $\{\beta_1^k\}$ 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 c is omitted from the equation. $\{\beta_1^k\}$ are measured relative to $\beta_1^{k=-1}$ which is omitted. For a given campaign event c and the campaign county $\{j\}_c$, $Connected_i$ is a binary variable indicating if county i has a above-state-median social connectedness with the campaign county. t^* is the month when the geographically distant campaign occurs. The analysis sample consists of only counties that are at least 750 miles away from the campaign county. 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: Distant Floods and Local Flood Insurance Purchases

This table shows results from the event study design: $Y_{it}^f = \beta_0 + \beta_1 * Connected_i^f \times PostFlood_t^f + \beta_2 * Connected_i^f + \beta_3 * PostFlood_t^f + \epsilon_{it}^f$. For a given flooding event f and the associated flooded 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 flooded 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 flooded area. Three different choices of distance threshold are considered: 500, 750, and 1,000 miles. $PostFlood_t$ is a binary variable indicating post-flood periods. 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. 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			
$Connected \times PostFlood$	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
R-squared	0.066	0.046	0.068	0.063
	Panel B: Restricted Sample Period 2014-2017			
$Connected \times PostFlood$	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
R-squared	0.060	0.035	0.061	0.058
Connectedness Weight	PW	EW	PW	PW
Distance Threshold	750 miles	750 miles	500 miles	1000 miles

Table 3: Heterogeneous Effects of Distant Floods

This table shows results from the event study design: $Y_{it}^f = \beta_0 + \beta_1 * Connected_i^f \times PostFlood_t^f + \beta_2 * Connected_i^f + \beta_3 * PostFlood_t^f + \epsilon_{it}^f$. Panel A and B consider two deviations from the baseline specifications presented in Table 2. All variables are defined as per Table 2 except that in Panel A, $Connected_i$ is defined as a binary variable that equals 1 if county i has a connectedness measure in the top quartile of the state and equals 0 if county i has a connectedness measure in the bottom quartile of the state; Panel B only includes the 18 significant flooding events defined by 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			
$Connected \times PostFlood$	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
R-squared	0.14	0.09	0.14	0.13
	Panel B: Subsample of Significant Floods			
$Connected \times PostFlood$	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
R-squared	0.09	0.04	0.10	0.08
Connectedness Weight	PW	EW	PW	PW
Distance Threshold	750 miles	750 miles	500 miles	1000 miles

Table 4: Spillover Effects Across Insurance Products

This table shows results from the event study design as described in Table 2. Panel A and B construct the analysis around flooding events and examine changes in health and earthquake insurance purchases: $Y_{it}^{f} = \beta_0 + \beta_1 * Connected_i^f \times PostFlood_t^f + \beta_2 * Connected_i^f + \beta_3 * PostFlood_t^f + \epsilon_{it}^f$. In Panel A, the dependent variable Y_{it} measures the health insurance coverage (in percent) for county i in year t. In Panel B, Y_{it} measures the number of earthquake insurance policies in-force (with January 2010 normalized to 100) in county i in year t (earthquake insurance data is only available for Missouri counties). Panel C considers a placebo test, which constructs the event study around non-flood natural disasters: $Y_{it}^{f} = \beta_0 + \beta_1 * Connected_i^f \times PostDisaster_t^f + \beta_2 * Connected_i^f + \beta_3 * PostDisaster_t^f + \epsilon_{it}^f$. In Panel C, Y_{it} measures the flood insurance demand in county i in month t, same as in 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)
	Panel A: Dependent Variable = Health Insurance			
$Connected \times PostFlood$	-0.012	0.009	-0.011	-0.015
	(0.016)	(0.016)	(0.014)	(0.020)
Observations	2,527,863	2,527,863	3,440,817	1,671,286
R-squared	0.013	0.010	0.013	0.013
	Panel B: Dependent Variable = Earthquake Insurance			
$Connected \times PostFlood$	0.096	0.247	0.134	0.502
	(0.460)	(0.435)	(0.407)	(0.523)
Observations	81,729	81,729	126,783	44,778
R-squared	0.005	0.007	0.003	0.010
	Panel C: Placebo Using Non-flood Disasters			
$Connected \times PostDisaster$	-0.115	-0.340	-0.055	0.217
	(0.277)	(0.277)	(0.247)	(0.307)
Observations	13,850,539	13,850,539	17,493,201	9,835,717
R-squared	0.041	0.036	0.033	0.035
Connectedness Weight	PW	EW	PW	PW
Distance Threshold	750 miles	750 miles	500 miles	1000 miles

Table 5: Alternative Methodology of Estimating the Causal Effect of Social Interactions

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

Table 6: The Intended Effects of Flood-risk-map Campaigns

This table shows results from the two-way fixed-effect regression: $Y_{it} = \alpha_i + \lambda_t + \beta * Campaign_{it} + X_{it} + \gamma * Campaign_{it} \times HighSCI_i + \epsilon_{it}$. 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 $Campaign_{it}$ is a binary variable indicating if county i has publicized its new flood risk maps at time t; this term is set to zero for the control counties without campaigns. α_i and λ_t are the county and year-month fixed effects. $HighSCI_i$ is a binary variable that equals one if county i has an above-median value of the intra-county relative probability of friendship. X_{it} are the covariates: $Premium_{it}$ is the average premium per policy (in \$); $Coverage_{it}$ is the average coverage per policy (in \$k); $Flood_{it}$ is the number of flood occurrences. 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)
$Campaign$	21.394***	21.266***	20.570***	5.922*
	(5.661)	(5.668)	(5.143)	(3.183)
$Premium$		-0.007	-0.005	-0.007
		(0.007)	(0.013)	(0.007)
$Coverage$		-0.058	0.005	-0.059
		(0.049)	(0.103)	(0.050)
$Flood$		0.987*	1.472	1.096*
		(0.553)	(0.905)	(0.566)
$Campaign \times HighSCI$				36.334***
				(12.349)
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

Table 7: Distant Campaigns and Local Flood Insurance Purchases

This table shows results from the event study design: $Y_{it}^c = \beta_0 + \beta_1 * Connected_i^c \times PostCampaign_t^c + \beta_2 * Connected_i^c + \beta_3 * PostCampaign_t^c + \epsilon_{it}^c$. For a given campaign event c and the associated campaign county $\{j\}_c$, county i's social connectedness to $\{j\}_c$ is measured by the relative probability of Facebook friendship $p_{i,j}$ obtained from Bailey et al. (2018b). $Connected_i$ is a binary variable indicating if county i is socially connected with the campaign county, which is defined as having a value of $p_{i,j}$ above the state-median. The analysis sample consists of only counties that are geographically distant to the campaign county. Three different choices of distance threshold are considered: 500, 750 and 1,000 miles. $PostCampaign_t$ is a binary variable indicating post-campaign periods. The dependent variable Y_{it} measures the insurance demand in county i in month t, which is 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)
$Connected \times PostCampaign$	1.21***	0.93***	1.30***	1.58***
	(0.18)	(0.11)	(0.16)	(0.22)
Observations	56,017,387	107,644,567	83,192,216	34,073,468
R-squared	0.03	0.03	0.03	0.03
Distance Threshold	750 miles	250 miles	500 miles	1000 miles

Appendix

A. Multiple Publications of the Flood-risk Maps

FEMA's Community Status Information (CSI) database only provides the publication date of the latest map. As FEMA aims to review its maps every five years, a county, in principle, may have two publication dates during my 10-year sample period. In this section, I present a set of evidence to show that FEMA does not achieve this goal; new publications are actually much more than five years apart.

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 is behind schedule.

Second, according to FEMA's Community Status Information (as of June 2020), almost 75% of the communities have an effective date more than five years old; that is, the latest update was before June 2015. Moreover, 37% (13%) of the maps are more than 10 (more than 20) years old. These statistics indicate that FEMA has not kept 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; that is, the communities with different dates in the two downloads. I find 412 such cases, for which I can impute the time spell between the two publications. It takes, on average, 11.5 years to publish a new map.

B. Survey of Facebook Posts

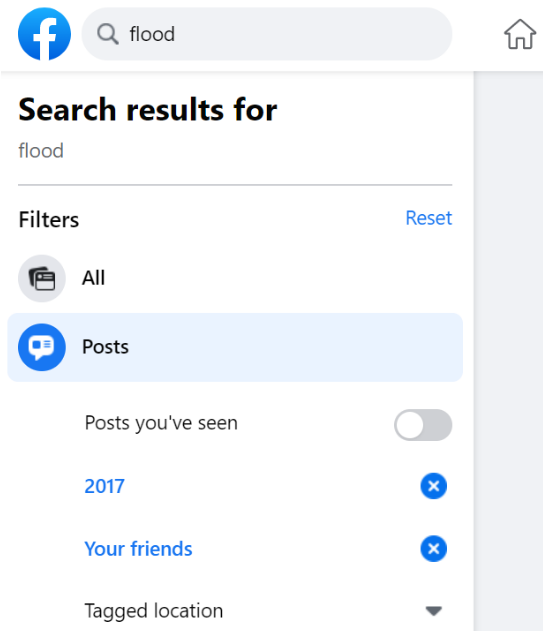
I conduct a survey of Facebook posts about flood risk and flood insurance to provide additional findings to support my hypothesis. As the social connectedness data described in Section 2.2 is aggregated and does not disclose information about account-level posts, I ask survey respondents questions that describe what information their friends share on Facebook during flooding events.

The survey is hosted on SurveyMonkey, one of the largest online survey platforms. I use SurveyMonkey's Audience panel to target specific and representative audiences.[23] In particular, I focus on residents in Texas and Louisiana, the two states most affected by Hurricane Harvey in August 2017.

The questionnaire is enclosed:

Please search keyword "flood" on your Facebook (as in the screenshot below) and answer the following 2 questions.

* set *Date posted* as **2017** and *Posts from* as **Your friends**



Q1: During the period of Aug–Dec 2017 (after Hurricane Harvey), how many posts (including shared posts) are about flood/flooding (e.g., texts/pictures/videos about damage)?

[23]Respondents are compensated for completing a survey. SurveyMonkey charged £2.38 per completed response for this project, but what fraction is received by the respondent is unknown.

Q2: During the same period, how many posts are about flood insurance?

[End of the survey]

As presented above, each survey contains two questions with instructions on how to search flood-related posts made by Facebook friends during a specific period. I am thus able to quantify and compare the informational content (flood experiences versus flood insurance) of Facebook posts during a major flood. I obtained 128 valid responses that provide numerical answers to both questions. Demographics are balanced according to censuses by SurveyMonkey Audience. Appendix Figure A.4 presents the distribution of gender, age, income, and respondents' electronic device.

Appendix Figure A.5 shows that, for the 128 responses received, the average number of flood-related posts is 13.1 (S.D.=25.5) while the average number of insurance-related posts is 2.6 (S.D.=4.2); the difference is statistically significant at the 5% level. Moreover, because individuals may differ in number of Facebook friends, I create a scale-insensitive measure by calculating the ratio of the two answers for each respondent. The third bar of Appendix Figure A.5 shows that the average ratio is 3.8 (S.D.=5.2), which is significantly larger than 1; this suggests that the number of flood-related posts is about three times higher than insurance-related posts.

C. Bias of Two-way Fixed-effects Estimators

It is now well-known in the applied econometrics literature that two-way fixed-effects (TWFE) estimations of difference-in-differences coefficients can lead to substantial biases when there are staggered treatment timing and heterogeneous/time-varying treatment effects (Baker et al., 2021; Borusyak et al., 2021; Callaway and Sant'Anna, 2021; De Chaisemartin and d'Haultfoeuille, 2020; Goodman-Bacon, 2021; Sun and Abraham, 2021).

In my application, Regression (5) described in Section 5.1 is a prototype for this type of setting that may feature the aforementioned biases: exploiting its staggered rollout, I estimate the effect of the NFIP's advertising campaign on local targeted households' flood insurance take-up using a TWFE model. I therefore follow a number of suggestions from the literature to verify the validity of my results.

Goodman-Bacon (2021) shows that the TWFE estimator is a weighted average of all possible 2×2 difference-in-differences (DD) estimator that compares timing groups to each other (referred to as the DD decomposition). Goodman-Bacon (2021) illustrates how to use the Stata command `bacondecomp` to decompose a TWFE estimator by plotting each 2×2 DD estimator against its weight, which provides diagnostic checks to assess if the average DDs vary across types of comparisons (in particular, if the comparisons to "already treated" groups are very different); hence, it gives a sense of how biased the TWFE estimator is.

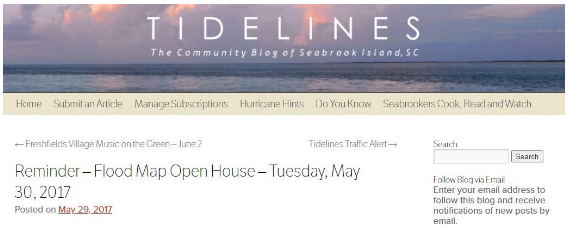
I try to implement the Goodman-Bacon decomposition `bacondecomp` in my application, but encounter practical challenges. In Goodman-Bacon (2021), the replication example is from Stevenson and Wolfers (2006), which exploits a staggered adoption of unilateral divorce laws in 37 states from 1969 to 1985. In contrast, my setting uses a panel dataset of 3,053 counties over 116 months. This creates issues when running `bacondecomp` in Stata, since the decomposition needs to compute all possible 2×2 DD estimators, exceeding the upper limit of "matsize" in Stata.

To overcome this problem, my solution is to reduce my sample size. Specifically, out of the 3,053 counties, each time I draw a random sample of 1,000 counties for which I am able to run bacondecomp and store the decomposition outcomes; I iterate this process 100 times. Appendix Figure A.6 below depicts all 2×2 DD estimators against their weights in one particular iteration. The average outcomes from the 100 iterations are presented in Appendix Table A.1. The results show that across the four types of comparisons, the average treatment effects are all positive and have fairly comparable magnitudes, suggesting that the bias of the TWFE estimate is quite small.

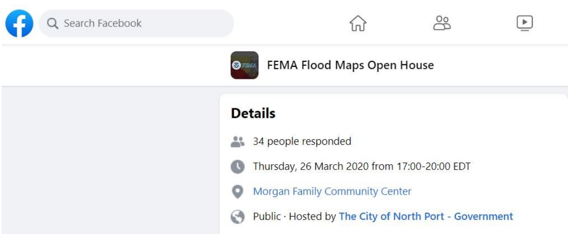
I also consider two other alternative estimators—proposed by Borusyak et al. (2021) (Stata command: did_imputation) and De Chaisemartin and d'Haultfoeuille (2020) (Stata command: did_multiplegt)—to address the bias from treatment effect heterogeneity in TWFE regressions. Appendix Figure A.7 presents the results using these two approaches; they are similar to those of the canonical TWFE estimators presented in the main paper (Figure 7), suggesting that the biases in my empirical setting are likely small.

Appendix Figures and Tables

(a) Open House Invitation on Local News



(b) Advertisement on Community Blog



(c) Open House Invitation via Facebook


November 17, 2016

ORDER PRINTS SHARE

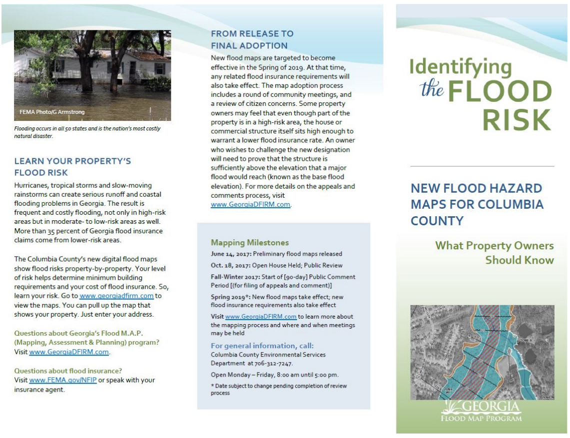
New Flood Maps Become Final in Willacy County, Texas

Targeted News Service (Press Releases)

DENTON, Texas, Nov. 16 -- The U.S. Department of Homeland Security's Federal Emergency Management Agency issued the following news release:

New flood maps become effective in Willacy County on April 5, 2017. The maps include the cities of Lyford, Raymondville and San Perlita, and unincorporated areas of Willacy County. Residents are encouraged to examine them so they can determine their need to buy flood insurance. By knowing their risks, individuals and community leaders can make informed decisions about building and development.

(d) Announcement of New Maps Publication



(e) Brochure

(f) Local Newspaper

Figure A.1. 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.

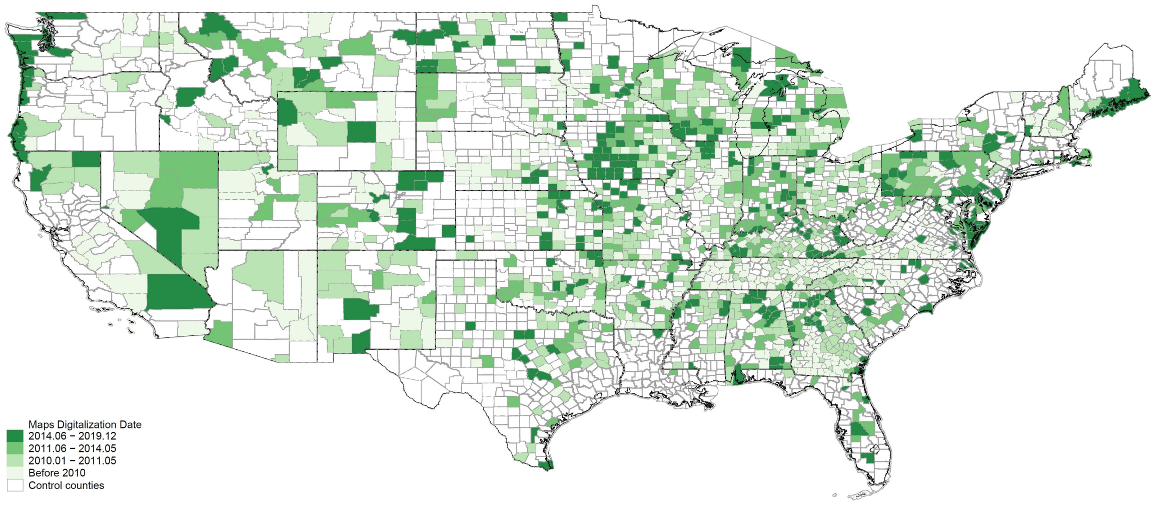


Figure A.2. Staggered Flood-risk-map Campaign across Counties and Time

The figure shows the flood-risk-map campaign by county and time. The darker shade represents the more recent publication date of the new maps. The unshaded counties represent the untreated group.

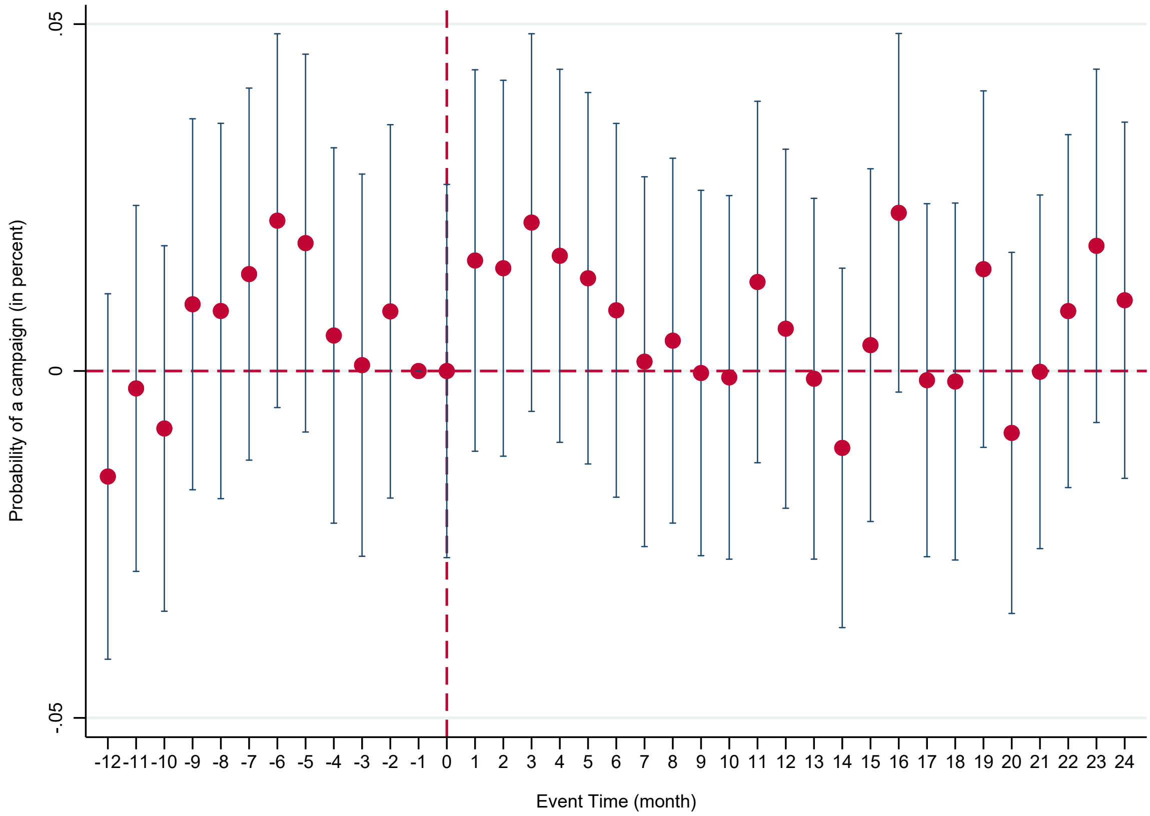
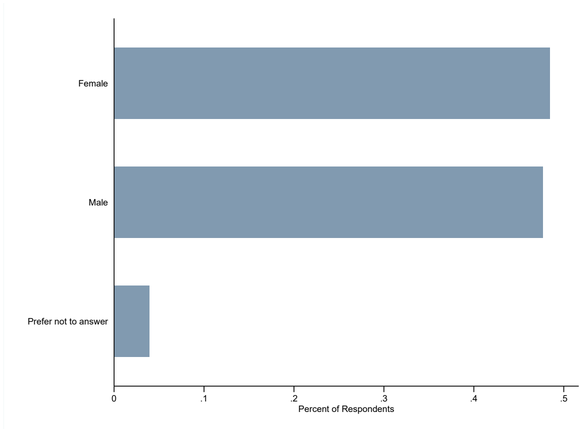
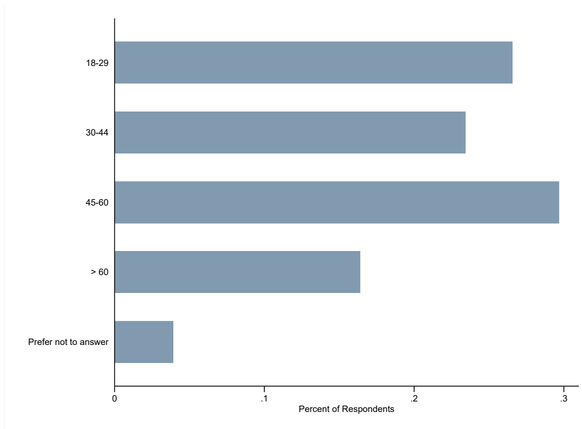


Figure A.3. The Difference between the Probabilities of Having a Campaign in the Treated and Control Counties

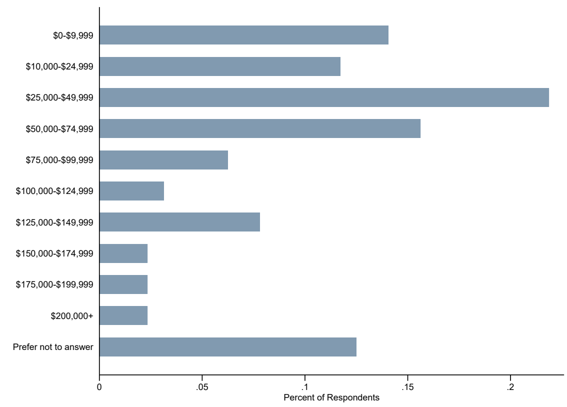
This figure plots the coefficient estimates of $\{\beta_1^k\}$ from the event study design: $Y_{it} = \beta_0 + \sum_k \beta_1^k * Connected_i \times \mathbb{1}(t = t^* + k) + \beta_2 * Connected_i + \sum_k \beta_3^k * \mathbb{1}(t = t^* + k) + \epsilon_{it}$. For notational brevity, the event index c is omitted from the equation. $\{\beta_1^k\}$ are measured relative to $\beta_1^{k=-1}$ which is omitted. For a given campaign event c and the associated campaign county $\{j\}_c$, $Connected_i$ is a binary variable indicating if county i has a above-state-median social connectedness with the campaign county. t^* is the month when the geographically distant campaign occurs. The analysis sample consists of only counties that are at least 750 miles away from the campaign county. The dependent variable Y_{it} is a dummy variable indicating if county i had the campaign in month t. Standard errors are clustered at the county level. The bands around the coefficient estimates show the 95% confidence intervals.



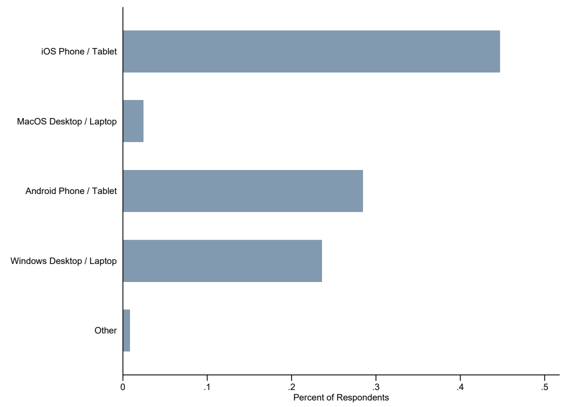
(a) Gender



(b) Age



(c) Income



(d) Device

Figure A.4. Demographics of Survey Respondents

This figure depicts the demographics of the survey respondents. Figure (a) to (d) show the distribution of gender, age, income, and the electronic device that the respondents used to complete the survey.

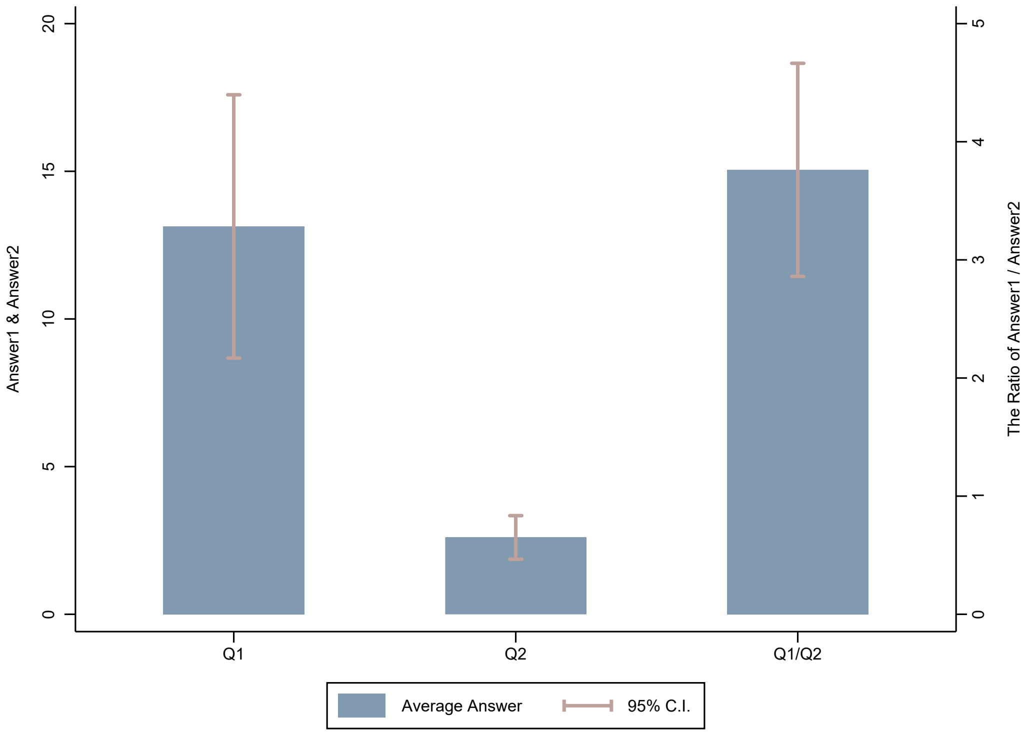


Figure A.5. Survey Responses

This figure depicts the responses to survey questions administered to a sample of individuals living in Texas and Louisiana, the two states that were most affected by Hurricane Harvey in 2017. Results are obtained for 128 respondents of the survey. The first two bars (corresponding to the left y-axis) show the average answers to questions "Q1: During the period of Aug–Dec 2017 (after Hurricane Harvey), how many posts (including shared posts) are about flood/flooding (e.g., texts/pictures/videos about damage)?" and "Q2: During the same period, how many posts are about flood insurance?" The third bar (corresponding to the right y-axis) presents the average value for the ratio of each respondent's two answers.

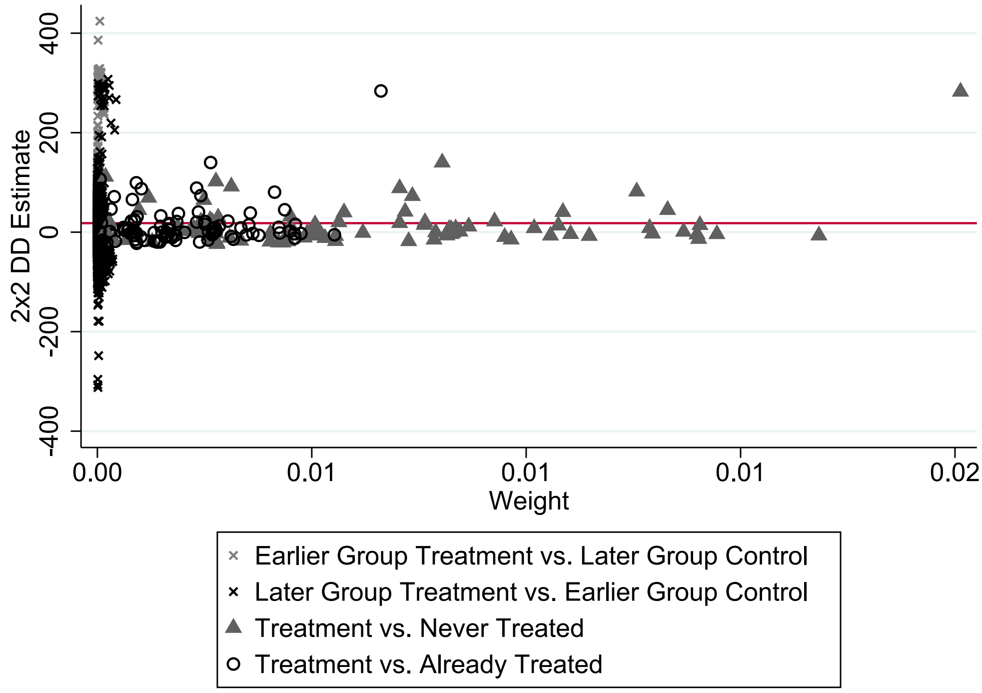
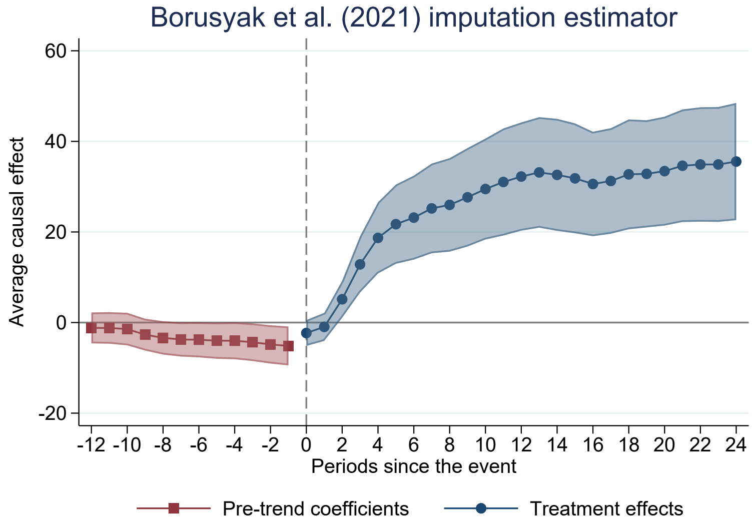
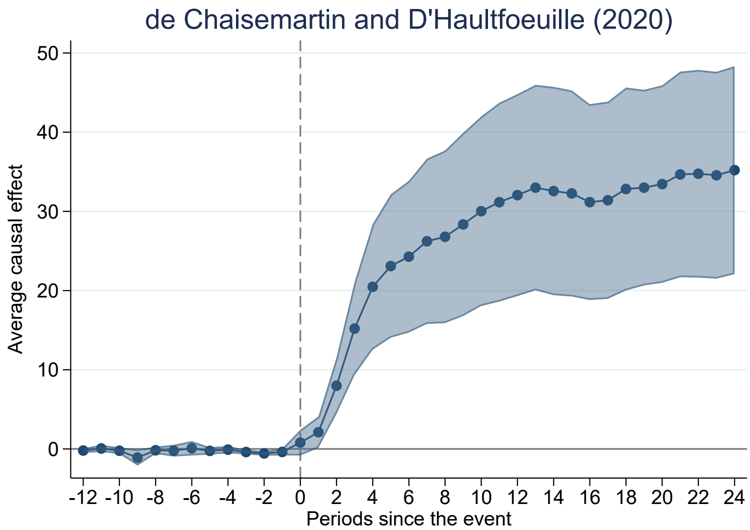


Figure A.6. The Goodman-Bacon Decomposition

Goodman-Bacon (2021) shows that the TWFE estimator is a weighted average of all possible 2×2 difference-in-differences (DD) estimator that compares timing groups to each other. This figure depicts the Goodman-Bacon (2021) decomposition, which plots each 2×2 DD estimator against its weight. Due to computational constraints in Stata (the upper limit of “matsize”), I am not able to run the command `bacondecomp` for my full sample (3,056 counties). This figure is generated based on a random sample of 1,000 counties.



(a) Borusyak et al. (2021)



(b) De Chaisemartin and d'Haultfoeuille (2020)

Figure A.7. Robust TWFE Estimators

This figure presents the results using the approaches proposed by Borusyak et al. (2021) and De Chaisemartin and d'Haultfoeuille (2020) to correct the biases of TWFE estimations of difference-in-differences coefficients (Regression 5).

Table A.1: The Goodman-Bacon Decomposition

Out of the 3,053 counties in my application, each time I draw a random sample of 1,000 counties for which I am able to run the Goodman-Bacon (2021) decomposition (Stata command: `bacondecomp`) and store the results; I iterate this process 100 times. This table shows the average outcomes from the 100 iterations.

TWFE estimate: 21.58		
DD comparison	Weight	Avg DD Est
Earlier Treatment vs. Later Control	0.076	27.437
Later Treatment vs. Earlier Control	0.158	16.562
Treatment vs. Never treated	0.569	21.093
Treatment vs. Already treated	0.198	24.656