Carnegie Mellon University
Tepper School of Business

DISSERTATION

*Submitted in partial fulfillment of the requirements
for the degree of*

**DOCTOR OF PHILOSOPHY
ECONOMICS**

Titled

"Essays on Beliefs about Catastrophic Risks"

Presented by

Timothy Hyde

Accepted by

Nicholas Muller	**4/28/2022**
Chair: Prof. Nicholas Muller	**Date**

Approved by the Dean

Isabelle Bajeux	**4/29/2022**
Dean Isabelle Bajeux	**Date**

Essays on Beliefs about Catastrophic Risks

by

Timothy Hyde

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics

Tepper School of Business

Carnegie Mellon University

Dissertation committee:
Nicholas Muller, CMU
Saurabh Bhargava, CMU
Dennis Epple, CMU
Lala Ma, University of Kentucky

Acknowledgements

This dissertation would not have been possible without incredible support from people in the CMU community and beyond. First and foremost, immense thanks to Nick Muller and Saurabh Bhargava for their guidance, advice, and encouragement. Nick's ability to kick-start the brainstorming process helped bring so many of my half-formed ideas to fruition, and his relentless optimism about my work helped me persevere at the lowest points of my graduate study. Through Saurabh's close and attentive mentoring, I learned so much about the research process, and his willingness to have me collaborate on his projects provided invaluable hands-on experience. I also thank Carrie Hopkins, who coauthored my third chapter and whose work inspired many of the ideas in this dissertation. Thanks also to Dennis Epple and Lala Ma for their service on my dissertation committee, and their encouragement, guidance, R code recommendations, and constructive criticism.

I also acknowledge invaluable assistance, constructive comments, data facilitation, and suggestions from Jay Apt, Linda Babcock, James Best, Justin Gallagher, Kareem Haggag, Tim Houlihan, Alex Imas, Karam Kang, David Kelly, Jennifer Kirby, Rebecca Lessem, Yanjun Liao, George Loewenstein, Stephen Spear, Lowell Taylor, Margaret Walls, Randall Walsh, and many more. In addition, Stephanie Rifai, Cassandra Taylor, and Sana Lakdawala provided excellent research assistance and data management.

Finally, thanks to my partner Micol and my son Milo, who have shown incredible patience and grace during this journey.

for Milo: may you enjoy a lifetime of learning

Contents

List of Figures

List of Tables

Chapter 1

How do homebuyers adapt after experiencing a natural disaster? Evidence from the Florida real estate market

1.1 Introduction

One of the biggest questions surrounding the issue of climate change is adaptation: will people adapt to rising temperatures and extreme weather events, how costly will these adaptations be, and will they be made before it is too late? There is an enormous literature studying climate adaptation around the world (Dell, Jones and Olken 2014; Mendelsohn, Nordhaus and Shaw 1994; Deschênes and Greenstone 2007; Kaiser et al. 1993). A key ingredient in adaptation is the initial understanding that a change needs to be made. Researchers have shown that first-hand experience with factors like excessive heat and natural disasters can promote belief in climate change (Deryugina 2013; Konisky, Hughes and Kaylor 2016) and spur adaptation in the context of agricultural investments or flood insurance takeup (Mase, Gramig and Prokopy 2017; Gallagher 2014). However, there is still relatively little empirical evidence on how exposure to climate change affects real estate purchases (Kahn 2016; Bunten and Kahn 2014; Boustan, Kahn and Rhode 2012). Any such response has major implications for coastal areas. Adverse demand shocks concentrated in these markets would destroy personal and commercial wealth, thereby shrinking municipal tax bases and reducing funding for schools and other public goods.

In this study, I use the timing and geographic extent of flooding events as an exogenous informational and psychological treatment on the people living nearby, and measure the effects of this "disaster shock" on subsequent participation in non-local real estate markets. I aim to shed light on two important questions: how do people translate experience with disasters into adaptive behavior, and how destructive are these adaptations for coastal real estate markets? The answers to these questions bear on whether coastal markets are overbuilt with properties that will likely be submerged within a few generations.

To analyze adaptive behavior after disasters, I pair data on over 4 million property sales from a database of Florida real estate transactions with information on over 400 flood disaster events across the United States. By focusing on the behavior of buyers who participate in Florida markets but hail from other states, I disentangle the psychological impact of disaster exposure from the direct physical damage impacts that can roil local real estate markets (Graff Zivin, Liao and Panassie 2020). Florida is a natural candidate for study as the state with the most real estate at risk from sea level rise over the next century (Union of Concerned Scientists 2018), as well as the most popular destination for out-of-state home buyers (Kerns and Locklear 2019).

I use the random variation in the timing of flood disaster declarations to identify the causal effect of recent flood exposure on home purchases in coastal Florida markets, and purchases of homes near the water in particular. I find significant evidence of retreat from the coast among buyers coming from counties with recent flood disasters, with shocked buyers purchasing 20-30% fewer homes within 1 km of the water (relative to farther inland) than under the counterfactual where they are not shocked. I use a random-utility sorting model to estimate how much shocked buyers devalue flood-prone properties, and find suggestive evidence that shocked buyers value properties in flood zones $500 -$3,000 less than comparable non-shocked buyers. I then employ hedonic price models to quantify the impact of disaster-shocked buyers on the marginal implicit price of coastal proximity. I find that a large influx of shocked buyers to a Florida market can significantly erode the amenity value of access to the water, reducing property values in the most affected parts of the market by 1-2%.

In Section 1.2, I survey the fast-growing literature on questions of climate belief, adaptation, and real estate markets. In Section 1.3, I describe ZTRAX, the real estate database, as well as other public data sources that allow me to complete this analysis. I then proceed to attack the question of adaptive behavior in this context from three directions. In Section 1.4, I introduce simple transaction count models that allow me to identify retreat behavior, and show significant and robust evidence of retreat in Florida markets among shocked buyers. In Section 1.5, I introduce a residential sorting model that I estimate in an attempt to value the different preferences of shocked buyers in dollars terms, and measure the welfare gain to shocked buyers of the information imparted by their past storm experience. In Section 1.6, I introduce a "meta" hedonic price approach that I argue allows me to estimate the threat that retreat behavior poses to the value of the coastal amenity currently capitalized into Florida homes. In Section 1.7, I discuss the implications of my findings and propose avenues for future research.

1.2 Literature review

Below, I survey the literature on the impacts of disaster exposure at both the individual and housing market level, along with past attempts to document adaptive behavior in the wake of disasters. This study contributes to this literature by demonstrating the transfer of disaster shock in one location to markets in other locations, and also identifying a plausible climate adaptation strategy – retreat from the water – undertaken by these shocked home buyers. To my knowledge, this study is the first to analyze the differential behavior of home buyers in the same market who have, and have not, recently experienced informational disaster shocks.

1.2.1 Belief updating after a disaster shock

There is a robust literature focused on measuring belief updating in the wake of weather events like excessive heat or natural disaster that implicate climate change. Konisky, Hughes and Kaylor (2016) and Shao and Goidel (2016) measure the effect of extreme weather events on climate attitudes, combining historical disaster data with survey data from opinion polls. Both studies find some effect of past local experience on concerns about climate change, while the latter finds that political affiliations are still a much more important determinant of beliefs. Konisky, Hughes and Kaylor (2016), in particular, find that the effect is only significant for extreme weather that occurred in the four months prior to the survey, and that less recent disasters have no statistically significant link with attitudes. Botzen, Kunreuther and Michel-Kerjan (2015) conduct a case study in New York City in the years after the disastrous Hurricane Sandy flooding in 2012. They find that people tend to overestimate likelihood of flood, and underestimate damage conditional on flood, but do not have the analogous before-flood data needed to attribute this to Sandy in particular. Some researchers have studied the link between more mild climate shocks, like increased temperatures, and climate opinion. Egan and Mullin (2012) and Deryugina (2013) both find that exposure to abnormally high temperatures in one's local area leads to increased belief that climate change is happening or will happen. Finally, in Chapter 3 of this work, we show that in the area affected by torrential flooding due to Hurricane Harvey in 2017, pessimism about flood risk faded over time in the wake of the disaster.

Another strand of the literature has tried to infer changes in beliefs from economic decisions. Gallagher (2014) finds that flood insurance uptake increases markedly in counties affected by flood disaster declarations and remains elevated for several years. Dessaint and Matray (2017) study the behavior of corporate managers, who tend to adopt more conservative portfolios and increase cash holdings when hurricanes strike nearby. Similarly, Cameron and Shah (2015) find that survivors of floods and earthquakes in Indonesia perceive higher general levels of risk and behave in a more risk-averse manner. Gibson, Mullins and Hill (2019) look at changing real estate markets in New York City to infer belief changes attributable to both Sandy and flood zone reclassification. However, neither is able to connect these responses to explicit belief measures. The present study fits most closely into this latter literature, as I am not able to measure changes in beliefs directly, but observe substantial changes in investment decisions (specifically house purchases) that are directly attributable to disaster shocks.

1.2.2 Disaster shocks in coastal real estate markets

Several studies have documented and quantified the effects of local natural disasters, but few have been able to decouple local flood damage from possible changes in beliefs, and only one has taken advantage of information on heterogeneous contemporary buyers to explain purchase decisions in the context of climate change.

The New York City real estate market in the wake of Hurricane Sandy has attracted particular researcher interest. Gibson, Mullins and Hill (2019) find that the publication of new flood maps and the arrival of Hurricane Sandy each separately had large depressive effects on real estate prices in flood-vulnerable parts of New York. Similarly, Ortega and Taspinar (2018) find that property values for non-damaged structures in areas affected by Hurricane Sandy declined 8% by 2017, five years after the storm. These papers are closely related to this study, in that they measure the effect of disaster shocks on real estate markets, but they can not disentangle the psychological or informational impact of the disaster from the physical damage to New York City neighborhoods.

A host of studies of other specific disasters suffer from the same limitation. Studies of the 1993 Mississippi River floods in St. Louis (Kousky 2010), a major flood in Georgia in 1994 (Atreya and Czajkowski 2014) and repeated flooding in Fargo, North Dakota (Zhang and Leonard 2019), as well as a series of hurricanes in Florida (Graff Zivin, Liao and Panassie 2020), all find significant negative effects of disasters on property values, including differentially worse effects for the most flood-prone properties. McCoy and Walsh (2018) find a similar result in the context of wildfires at the wildland-urban interface near Denver, Colorado. However, a common limitation of all these studies is that they cannot separate the psychological and informational impact of disasters on the people living there from the impact on the physical infrastructure in the market. The present study surmounts this problem by focusing on residents who are shocked by an event in one location, and then enter a real estate market in Florida, far from the original disaster.

1.2.3 Climate adaptation in real estate

Few studies have specifically addressed the question of how ongoing climate change will affect long-term migration patterns, as opposed to the immediate migrations induced by mass destruction (Kahn 2016). Albouy et al. (2016) use hedonic models of U.S. cities' climate characteristics to predict that excessive heating will drive more people to migrate to the Pacific Northwest, but do not attempt to predict how the real estate markets within metropolitan areas will be affected. Bernstein, Gustafson and Lewis (2019) use the same database as the present study to analyze homes at risk of sea level rise (SLR) and find that these homes sell for less than homes that are similar on observables. The effect is concentrated among second-home purchasers, who the authors infer are more sophisticated than other buyers based on the education and income levels in their home ZIP codes. Similarly, Baldauf, Garlappi and Yannelis (2018) find that the SLR discount is more pronounced in "believer" neighborhoods where more survey respondents indicate a belief in that climate change is happening. Keys and Mulder (2020) find that buyers in Florida real estate markets seem to have become more sensitive to sea-level-rise zones over the past two decades – although sellers' attitudes have not changed – and postulate that this could be due to rising awareness of climate change generally. Murfin and Spiegel (2020), by contrast, find that what they term "relative sea level rise," projections that account for land subsidence, is not properly capitalized by markets.

Another set of studies takes advantage of changes and updates to official flood maps that occur periodically, even in the absence of any local disaster. These studies detect responses in real estate markets to determine the extent to which flood risk information is capitalized in this market, generally finding that there is a response (Hino and Burke 2021; Mulder 2021; Shr and Zipp 2019).

I am aware of two studies that analyze the effect of non-local shocks that could plausibly have affected real estate market participants only through their information content. Muller and Hopkins (2019) study the New Jersey coastal real estate market and use non-local disasters like Hurricane Katrina that garnered national media attention to measure the role of information in coastal real estate markets, finding a significant negative effect on price for homes in flood zones. Hallstrom and Smith (2005) study the effect of Hurricane Andrew – a massively destructive storm that ravaged Miami-Dade County and other parts of southern Florida in 1992 – on nearby Lee County, which narrowly escaped damage. They also find that homes in flood zones lost significant value, which they attribute to buyers and sellers recognizing the information conveyed by the storm and adapting their market behavior.

In sum, these studies explore purchasing behavior in areas impacted by disasters, but few considerd how buying behavior differs across heterogeneous buyer types (Baldauf, Garlappi and Yannelis 2018; Keys and Mulder 2020). To my knowledge, no study examines the differential buying behavior of individuals who have experienced disasters far from a given real estate market. This geographic remove is crucial to establish the informational and psychological aspects of disaster exposure apart from the equilibrium effects of local disaster damage.

1.3 Data

To study the question of climate adaptation in real estate markets, I rely on comprehensive Florida real estate transaction and assessment data from ZTRAX, a database provided by Zillow Inc (Zillow 2021).[1] This database provides detailed information on each property transacted and, crucially, geographic characteristics of buyers including their designated mailing addresses, which in the basis for my assignment of experienced flood history to buyers.

I complement the ZTRAX database with hydrological data on each property from U.S. government sources. With precise latitude and longitude coordinates from ZTRAX, I match each property to a flood zone classification, a sea-level-rise submersion zone, and calculate a straight-line distance to the water. Geographic data on flood zone classifications as designated by the Army Corps of Engineers is available from the U.S. Federal Emergency Management Agency (2021), while sea-level-rise projections are available from the National Oceanic and Atmospheric Administration (2021). I use QGIS mapping software to calculate the distance from each property location to the water with a detailed shoreline map from the University of Florida GeoPlan Center (2019).

Additionally, I draw on the Florida Cooperative Land Cover map maintained by the Florida Fish and Wildlife Conservation Commission (Florida FWC 2021) to identify neighborhood amenities such as parks and recreational sites. Finally, to assign a treatment measure to each observation, I use county-level historical disaster data maintained by the U.S. Federal Emergency Management Agency (Federal Emergency Management Agency 2019).

In the remainder of this section, I describe these sources in more detail.

1.3.1 ZTRAX database

The ZTRAX database provides information on real estate transactions and assessments across more than 2,700 U.S. counties dating back to the 1990s. In this study, I restrict focus to data on properties sold in Florida between January 2005 and December 2017, the period for which the transaction data was the most comprehensive at the time I accessed it. Zillow compiles data from county authorities and claims essentially universal coverage of transactions during this period (Zillow 2021). In approximately 48% of cases, I can identify the buyer's county of origin, as indicated by the buyer's mailing address, and in nearly every case I can identify the precise geographic location of the property transacted. ZTRAX data contains many observations with missing or seemingly erroneous information. I am forced to drop a small percentage of observations without listed prices or with listed prices of $0. I also exclude properties with listed sale prices above $10 million, because it appears that some of these are likely data entry errors (these account for less than 0.5% of all transactions). Additionally, some properties are

[1] Data provided by Zillow through the Zillow Transaction and Assessment Dataset (ZTRAX). More information on accessing the data can be found at http://www.zillow.com/ztrax. The results and opinions are those of the author and do not reflect the position of Zillow Group.

Table 1.1 · Descriptive statistics

	(1) Full sample, 2005-2017 (n = 4,764,491)	(2) Matched to assessment data (n = 4,451,961)	(3) Known origin only (n = 2,273,503)	(4) Buyers from 818 main origin counties (n = 963,412)	(5) Buyers from main origin counties, frontage only (n = 84,634)
Property characteristics					
Sales Price ($k)	247	245	227	265	504
	(350)	(335)	(378)	(429)	(695)
Square footage	1,953	1,945	1,844	1,901	2,071
	(1,023)	(1,019)	(1,003)	(983)	(1,329)
Number of stories	1.2	1.2	1.18	1.17	1.27
	(0.48)	(0.48)	(0.46)	(0.5)	(0.57)
Year built/remodeled	1989	1989	1989	1993	1989
	(19)	(19)	(20)	(17)	(17)
Single-family home	61%	62%	53%	48%	37%
Has garage	30%	30%	25%	27%	21%
Has pool	19%	20%	17%	20%	32%
Has boatslip	2%	2%	2%	3%	27%
Hydrological characteristics					
Distance to water (km)	3.55	3.47	3.3	3.2	.02
	(4.86)	(4.65)	(4.84)	(4.63)	(.01)
Frontage (<50 m to water)	6%	6%	7%	9%	100%
100-year flood zone	19%	19%	22%	23%	67%
500-year flood zone	29%	29%	31%	32%	76%
6-ft SLR submersion zone	14%	14%	16%	15%	54%
Transaction characteristics					
Floreclosure sale	21%	21%	24%	20%	13%
Indeterminant origin county	52%	53%	0%	0%	0%
Buyer origin outside FL	16%	16%	33%	74%	83%
Months since last flood in buyer's county	122	122	118	104	104
	(54)	(54)	(59)	(69)	(74)
Floods in last ten years in buyer's county	3.33	3.34	3.21	2.99	2.76
	(2.06)	(2.07)	(2.01)	(1.92)	(1.81)

Mean values and standard deviations for key variables in the sample. Column 1: the full sample include all residential home purchases in Florida between January 2005 and December 2017 for which reliable sales price information, transaction details, and precise geographic location data are available. Column 2: this sample excludes transactions for which thorough assessment data was not available, internally contradictory, or otherwise not reliable. Column 3: a transaction is included in the "known origin only" sample if the buyer lists a correspondence address that is not identical to the property address. Column 4: a county is included in the 818-county group if there are at least 100 purchases in the full sample associated with that county. Column 5: a subset of the sample in Column 4, inluding only properties within 50 meters of the water. Source: Author's analysis of Zillow ZTRAX database, and GIS calculations based on University of Florida coastline maps, FEMA flood zone maps, and NOAA sea-level-rise maps.

Table 1.2 · Transactions by metropolitan statistical area (MSA), origin data only

Metropolitan area	# of Ctys	Sales count (k)	Mean sales price ($k)	Non-FL buyer	New York buyer	Commonest non-NY origin	Condo	Flood zone	Sea level rise zone
Daytona Beach	2	118	176	34%	7%	New Jersey	19%	12%	6%
Destin	2	66	389	53%	1%	Georgia	28%	10%	2%
Ft Myers	1	259	202	47%	6%	Illinois	23%	30%	8%
Homosassa	1	41	89	31%	6%	New Jersey	2%	9%	7%
Jacksonville	5	159	177	24%	2%	California	15%	10%	6%
Miami	3	676	269	22%	7%	New Jersey	48%	19%	9%
Naples	1	104	463	61%	7%	Illinois	50%	16%	10%
Panama City	2	45	223	53%	1%	Georgia	36%	13%	6%
Pensacola	2	50	149	28%	1%	Alabama	9%	12%	4%
Port St. Lucie	2	85	184	33%	8%	New Jersey	18%	5%	3%
Punta Gorda	1	72	130	47%	6%	Massachusetts	11%	44%	19%
Sarasota	2	152	271	47%	6%	Ohio	28%	18%	6%
Space Coast	1	84	146	31%	6%	California	19%	8%	4%
Tampa	4	365	167	27%	4%	California	21%	12%	6%
Vero Beach	1	32	275	40%	7%	New Jersey	22%	19%	8%
Full sample	30	2,307	228	33%	6%	New Jersey	30%	17%	7%

Source: Zillow ZTRAX database and U.S. Census metropolitan area definitions. Information on properties in flood zones and sea-level-rise zones is provided by FEMA and NOAA respectively. Figures include all transactions between January 2005 and December 2017 for which reliable sales price data and buyer origin data are available. As such, they reflect only 48% of total transactions during this period.

listed with latitude and longitude coordinates that are inconsistent with county or ZIP code information provided; these are dropped where they can be identified. See Appendix A1 for more details about the myriad reasons that some transactions had to be dropped from the main sample.

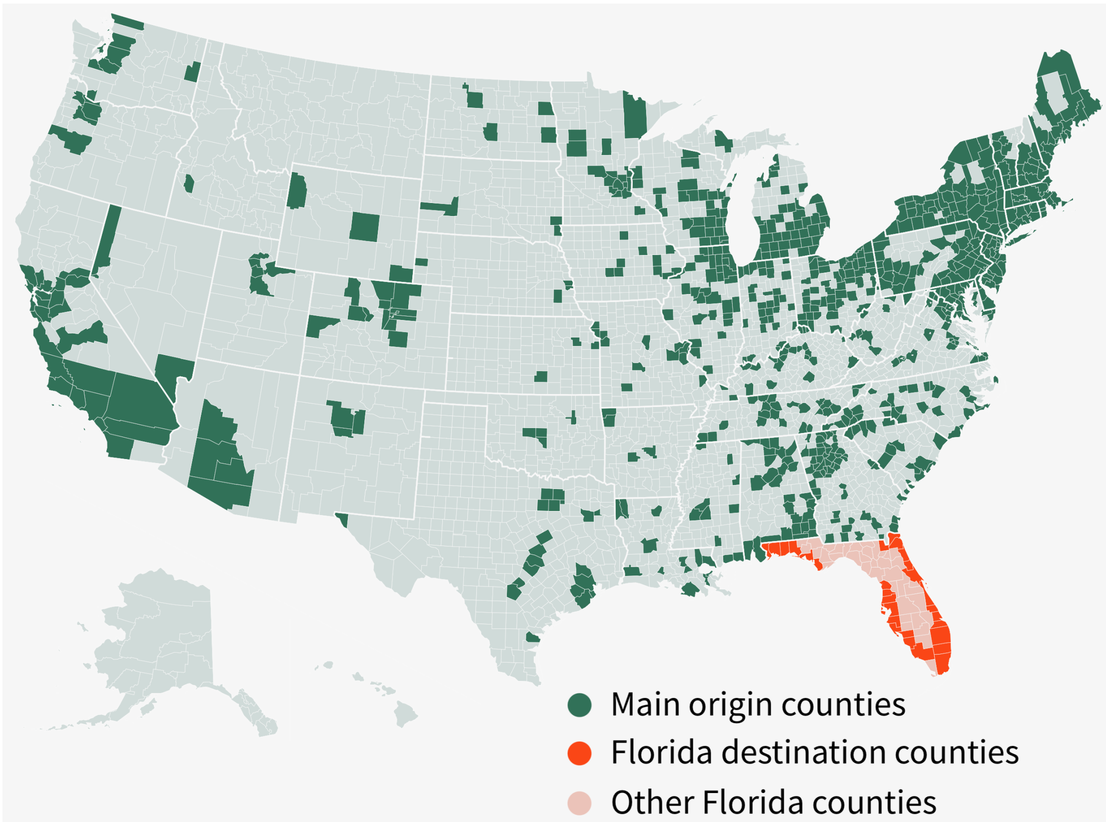
Condominium properties present an additional challenge for my analysis. People who live in condominiums or high-rise buildings are effectively decoupling access to the water from flood risk, which indeed may be part of the appeal. I cannot reliably observe what floor a property is located on in a large building in the ZTRAX data. In some analysis, I treat condos separately from other observations as the geocoding of condo properties to flood zones may not represent an accurate assessment of that property's risk, even abstracting away from any inaccuracies in the flood maps themselves. On a similar note, I cannot distinguish properties where the living areas are raised on stilts or otherwise elevated, leading to some inevitable measurement error in my flood risk measures.

ZTRAX assessment data drawn from government sources has a rich complement of structural variables for each property, although for about 17% of transactions I am unable to successfully match the underlying property to assessment data, usually due to a lack of assessment record for a particular parcel or incomplete records without key data on square footage or building age. Unmatched observations are still used in some parts of the analysis, but cannot be included in models that rely on property characteristics like square footage.

Table 1.1 displays descriptive statistics for the sample. The full sample (Column 1) include all residential home purchases in Florida between January 2005 and December 2017 for which reliable sales price information, transaction details, and precise geographic location data are available. Columns 2 and 3 show two distinct sub-samples of this full sample. In Column 2, the sample excludes transactions for which thorough assessment data were not available, internally contradictory, or otherwise not reliable. This is the core sample for the hedonic analysis in Section 1.6. A transaction is included in the "known origin only" sample (Column 3) if the buyer lists a correspondence address that is not identical to the property address (see Section 1.3.2 below for more on this issue). Column 4 presents a subset of the known-origin sample that includes only non-local buyers (i.e. not from the same MSA as the property purchased) and is restricted to the 818-county group with at least 100 purchases in the full sample associated with that county. This is the core sample used in the count models in Section 1.4. Column 5 displays a subset of the sample in Column 4, including only properties within 50 meters of the water,

Figure 1.1 · Map of 818 main Florida origin counties

Counties qualify as origin counties if there are at least 100 transactions in the sample between 2005 and 2017 where the buyer address is associated with that county. The counties comprising the 15 metropolitan areas in Florida included in the study are highlighted in orange.



which I term "frontage" properties.

The descriptive statistics reveal some distinctive features of homes purchased by out-of-MSA buyers (Column 4), relative to the broader group of known-origin homes in Column 3. While these homes are nearly 16% more expensive than the average home sold, they are less than 4% larger. They do, however, enjoy significantly better access to the water, being 4% closer to the water on average, and slightly more likely to be in the frontage zone (within 50 meters of the water) and the 100-year flood zone. These differences provide suggestive evidence that out-of-state buyers prioritize access to the water more highly than local buyers when making purchase decisions.

The housing market in the frontage zone, which may be most radically affected by changing beliefs about disaster risk, is distinctive from other parts of the market. Frontage homes are nearly twice as expensive as the average property, and are significantly larger, although slightly older on average. These deluxe properties likely occupy a separate market apart from the general coastal real estate market, which might explain the relatively resilient demand for these homes among shocked buyers that I find in some models.

Table 1.2 breaks down the sample by metropolitan area (MSA). I restrict my focus to these MSAs, which are the 15 coastal MSAs in Florida as defined by the U.S. Census Bureau, and contain over 95% of coastal county observations.[2] The MSAs range in size, average home price, and the share of buyers coming from out of state,

[2] A sixteenth Florida MSA, Tallahassee, includes two coastal counties, but I exclude it for two reasons. First, the coastal areas in this MSA are relatively sparsely populated, as they are mostly wetlands covered by National Wildlife Refuge status and other state and federal protections. Second, the central MSA area around Tallahassee, where most of the population lives, is not itself located near the coast. In all other cases, the coastal MSAs I do include are centered around cities that are themselves coastal.

reflecting the fact that some areas are more popular retirement or second-home markets, while others may be attracting mostly people relocating for work or family. Out-of-state buyers are most prevalent in Fort Myers, Sarasota, Naples, and Punta Gorda, which are clustered together on the Gulf Coast in the southwestern part of the state.

Figure 1.1 displays a map of the 818 counties that I include in my count models. Counties qualify as origin counties if there are at least 100 transactions in the main sample between 2005 and 2017 where the buyer address is associated with that county. Unsurprisingly, there is a noticeable skew toward the eastern half of the United States, particularly the most populous urban and suburban areas. Metropolitan areas in the northeast are especially over-represented. I limit myself to these counties to reduce the number of observations with zero counts in the negative binomial regressions. This group includes over 94% of out-of-state buyers in the nationwide sample.

1.3.2 Determining buyer origin

To apply my treatment variable, which is the amount of time since the most recent flood disaster in a buyer's home county, I need to reliably determine where each buyer was living in the months and years prior to the transaction. The ZTRAX data for each transaction contains information on the buyer's mailing address as recorded by the local government for future tax correspondence purposes. I use this variable to identify the buyer's home ZIP code and county, and in turn to define the treatment variable for each observation.

In cases where a buyer is purchasing a home in Florida and local officials record a residential address in Wisconsin or Massachusetts as the buyer's mailing address, it seems reasonable to assume that this buyer was previously living – and possibly is continuing to live – at the out-of-state address, and the new property is an investment or a second home.

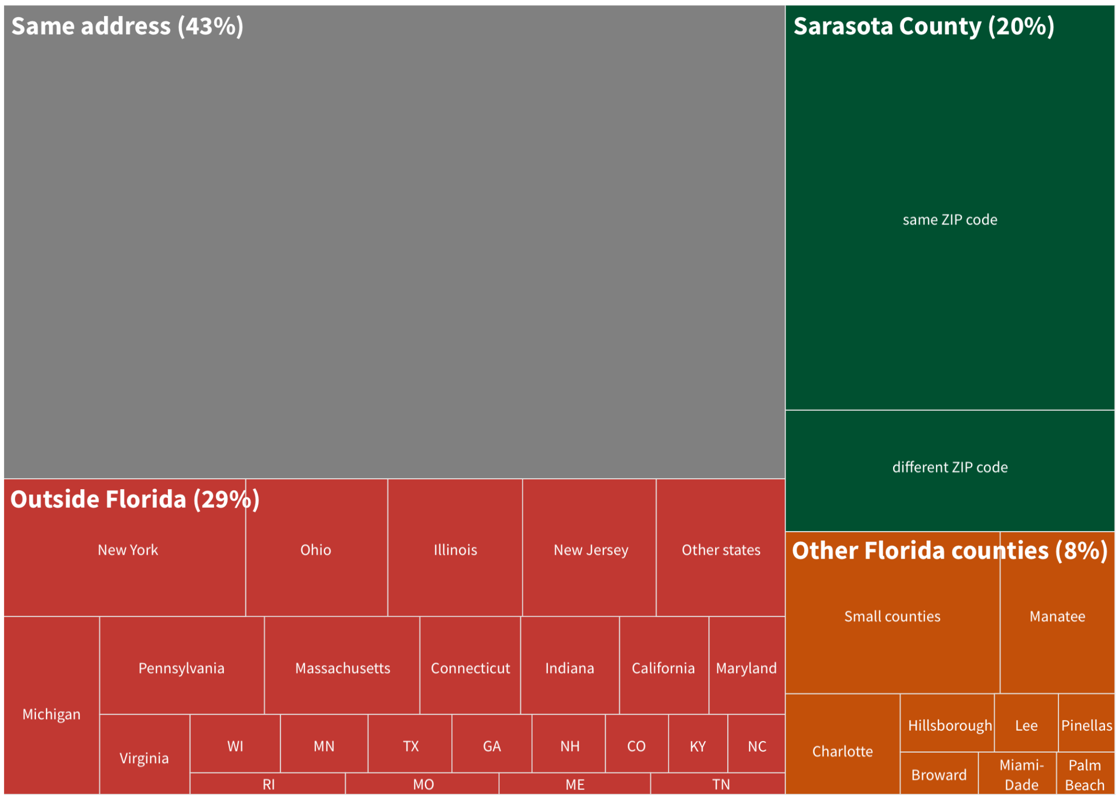
However, in approximately 52% of cases, the buyer's mailing address exactly matches the address of the property purchased, which I term the "same-address listing" phenomenon (see Figure 1.2 for an example breakdown of the buyer address variable in Sarasota County). In these cases, the buyer's address does not seem like a reasonable basis for determining a buyer's *previous* location. The buyer could plausibly have been previously living in Wisconsin, Massachusetts, some other part of Florida, or right down the street. In these cases, I have no way to identify the origin of these buyers, and no way to impute their recent flood history.

Previous literature has interpreted the same-address listing behavior as intent to immediately occupy the home, and to receive official correspondence at the new address going forward (Chinco and Mayer 2016; Graff Zivin, Liao and Panassie 2020). This would imply that non-same-address listers, by contrast, comprise some combination of investment companies, second-home purchasers, and people who intend to relocate to Florida permanently or semi-permanently in the intermediate future but not right away. Assessment data from ZTRAX includes information about whether properties were assessed as owner occupied and whether owners are claiming the homestead tax credit, which implies owner occupancy in most cases. These data sources are somewhat incomplete and mutually contradictory, and do not correspond closely to same-address listing behavior either.

Regardless of the root causes of this behavior, my analysis in Sections 1.4 and 1.5 will necessarily be limited to comparisons within the large minority of buyers where this address data does not match the property in question. I argue that this analysis is still valid and important. First, Florida is a major market for investment and second-home purchases, and this market is important in its own right. Second, previous research shows that non-owner-occupiers tend to be more savvy about climate risk (Bernstein, Gustafson and Lewis 2019), and detecting changes in their behavior in response to disaster experience may be of particular interest. Finally, I show evidence in Appendix A2 that while there is a statistically significant relationship between same-address listing behavior and incoming buyers' disaster experience in some MSAs, the economic significance of the relationship is small. For this relationship to explain my results would require a very peculiar selection effect: people from recently-shocked counties are no less likely to buy homes near the water, but those buying homes near the water are likelier to occupy them immediately.

Figure 1.2 · Tree map of "buyer mail" variable in Sarasota County transactions

A tree map of the buyer mailing address variable from ZTRAX transactions in Sarasota County, 2005-2017. Each rectangle represents a mutually exclusive group of transactions, with the size of the rectangle representing the group's relative size. Buyer mailing address is the key variable I use to determine each buyer's disaster shock treatment. Buyers who list a mailing address that matches the address of the property purchased ("same address" buyers) comprise a large minority of the sample in most counties, and their origins can not be identified. These purchases are excluded from portions of the analysis that rely on identifying buyers' origin counties. See text for more details about my interpretation of the same-address listing behavior.



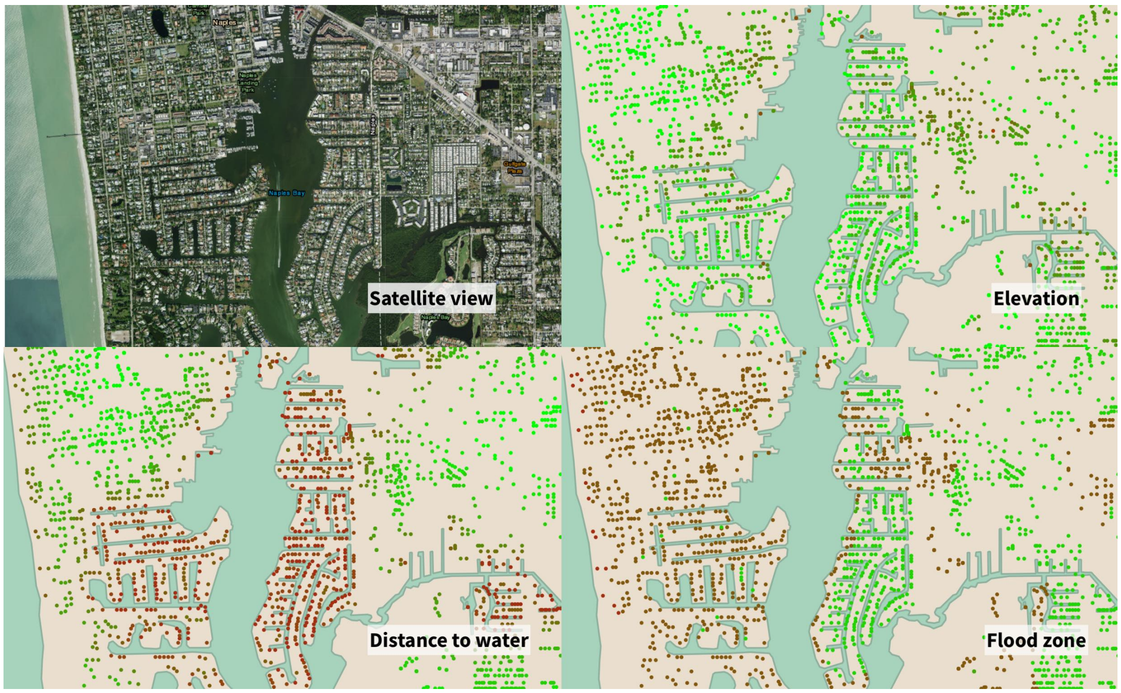
1.3.3 Hydrological data

I complement the ZTRAX database with data that pertain to each property's proximity to water, a factor which is generally understood as an attractive amenity, but which may also repel disaster-shocked buyers in particular. Using precise latitude and longitude information from ZTRAX, I match each property to a flood zone classification from the Army Corps of Engineers using flood maps made available by the U.S. Federal Emergency Management Agency. I categorize properties into five zones: A (corresponding to a 1% or greater annual risk of inundation),[3] VE (areas with 1% or greater risk and additional risk from storm surge flooding), AH (corresponding to a 1% or greater annual risk of shallow flooding or pooling), X (corresponding to an annual risk of inundation between 0.2% and 1%), and the "minimal flood hazard" zone, corresponding to a flood risk lesser than that of the 500-year flood zone. See Appendix Figure A1 for an example flood map as it appears on the FEMA website. Likewise, I match each property to one of seven feet-of-SLR submersion zones identified by National Oceanic and Atmospheric Administration (2021) that correspond to the amount of future sea level rise that would result in property submergence.

[3] The "A" flood zone designation has been replaced by "AE" on updated flood maps, but in either case they designate areas where FEMA is warranting a 1% annual risk of flood or greater.

Figure 1.3 · Hydrological characteristics of coastal properties in Naples, Florida

These screen captures from QGIS software illustrate properties from the sample superimposed on a highly-detailed map of the Florida coastline, including inlets, canals, and artificial water features like marinas. In each capture, properties are color-coded according to a different measure of flood vulnerability, with red properties facing the greatest risk and green the least. Quantities calculated using property latitude and longitude data mapped to flood zone data from FEMA, coastal boundary data from the University of Florida GeoPlan Center, and sea-level-rise data from NOAA.



Obtaining information on each parcel's distance to the water is more involved. I use GIS mapping software to calculate the straight-line distance from each property location to a detailed shoreline map from the University of Florida GeoPlan Center (University of Florida GeoPlan Center 2019). This allows me to observe homes that are relatively far from the open ocean but close to features like inlets, canals, and marinas that could be a source of floodwater, and code them accordingly (see Figure 1.3).

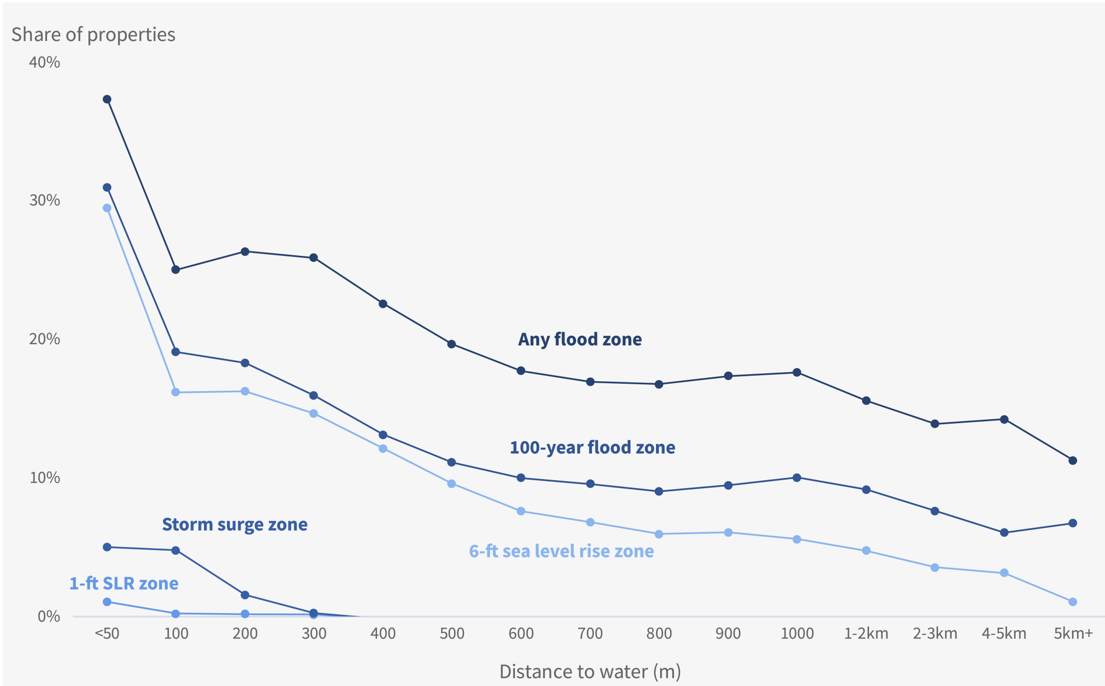
Figure 1.4 illustrates the distribution of flood zones for all properties in the 15-MSA sample within 10 km of the water. Owing to the state's famously flat topography, properties in Florida face significant flood risk even when they are relatively far from the water. While naturally a large percentage of properties in the "frontage" zone (50 m from the water or less) are in the 100-year flood plain, a non-negligible fraction of properties even further inland are also so labeled. For example, over 15% of homes between 2 and 4 km from the water are in one of the two flood zones. This has implications for *where* I would expect retreat behavior to arise, not simply from the immediate frontage areas to the areas a bit inland, but even from areas more than one kilometer from the water. Below, I show results indicating that retreat is detected in the region 50 m to 1 km from the water.

1.3.4 Flood event data

My source of identification in the analysis below is the random timing of major flood events in U.S. counties over the 2005-2017 period, as determined by Presidential Disaster Declarations (Gallagher 2014). I use historical county-level disaster data maintained by the U.S. Federal Emergency Management Agency. Each disaster is listed with a geographic range and an official description, such as "severe winter storms, heavy rains, and flooding." A

Figure 1.4 · Flood hazard by property distance to water

Share of transacted properties with each of five flood risk measures across 15 distance-to-water bins. Owing to the state's famously flat topography, many properties in Florida face significant flood risk even at a great distance from the coast. Quantities calculated using property latitude and longitude data mapped to flood zone data from FEMA, coastal boundary data from the University of Florida, and sea-level rise data from NOAA.



disaster was included in the sample if its description included the words "flood," "flooding," or "hurricane," or it was coded by FEMA as a coastal storm, dam/levee break, flood, hurricane, tsunami, or typhoon. This inclusive definition admits certain incidents that are categorized by FEMA as tornadoes or ice storms, but which include flood damage.

In the 2005-2017 period, there were approximately 430 distinct flood events in the United States,[4] resulting in 10,104 county-level disaster declarations that met the above definition. Table 1.3 displays descriptive statistics on the treatment variable. The average U.S. county experienced 0.28 such disasters per year during this 12-year period. This frequency underscores the fact that disaster declarations are not reserved for massive flood events, but include relatively less severe incidents as well. The treatment should therefore be considered a measure of recent exposure to moderate local flood events as opposed to exposure to historic, catastrophic flooding. In most cases, a single disaster event prompts disaster declarations over a wide area; the average disaster declaration includes nearly 30 counties. The most widespread disaster declaration during this period was for Hurricane Rita, a 2005 hurricane that caused flooding in multiple states and resulted in disaster declarations across 323 counties and county-equivalents in Texas, Louisiana, and Florida.

For each transaction in the ZTRAX database, I define treatment based on the time elapsed since the most recent flood event with these characteristics in the buyer's county of origin, as determined by the buyer mailing

[4]FEMA does not formally group disaster declarations across state lines. I group county-level flood declarations based on incident date, or storm name in the case of hurricanes, and find 430 separate "events." This likely provides an overestimate, as incidents beginning on consecutive dates in neighboring regions may pertain to the same weather system and be considered part of the same flood event by observers, but would be treated as separate disasters by my method. This grouping exercise does not affect the value of the treatment variable, as the relevant level for defining treatment is the county-month.

Table 1.3 · U.S. flood disaster frequency, 2005-2017

Total flood disaster events	**430**	**Transactions with known origin**	**2,273,503**
Incident type		*Distribution of treatment*	
Severe storm	69.7%	0-6 months	13.0%
Flood	20.5%	7-12 months	15.0%
Hurricane	4.7%	13-24 months	13.9%
Other	5.1%	25-36 months	8.9%
		37-48 months	6.7%
Counties per flood event	21.9	49-60 months	6.1%
County-level disasters	9,436	More than 5 years	36.5%
Mean annual disasters per county	0.297		
		Mean months since disaster:	50.3
Pareto stat: top 36% of counties have 64% of disasters		Median months since disaster:	34

Source: U.S. Federal Emergency Management Agency. A disaster was included in the sample if its description included the words "flood" or "hurricane," or it was coded by FEMA as a coastal storm, dam/levee break, flood, hurricane, tsunami, or typhoon. Flood disaster event count is approximate while county-level disaster count is not; see text.

address variable. Although I only study transactions in the 2005-2017 period, I account for disasters that occur as early as 2000 to accurately calculate the treatment variable for counties with infrequent storms. Table 1.3 lists the distribution of this treatment variable across transactions. Given the frequency of flood disaster declarations in these areas, about 28% of the sample has experienced a disaster in the last 12 months, and 37% of the sample has not experienced a disaster for more than five years. I take each of these groups as a separate treatment category, reserving the group without any recent floods as the control group in my count models.

Ideally, I would observe each individual buyer's personal residential history and apply the treatment accordingly. However, I only observe the most recent previous address of home buyers (see caveats in Section 1.3.2 above), and I cannot determine how long they were living at their previous addresses. Therefore, I do not know if a person who moved from County X and purchased a home in Florida in 2010 was living in County X during a flood event there in 2007. I merely know that this buyer lived in County X, or at least was maintaining a mailing address there, immediately before purchase in 2010. Hence, the treatment I define is a proxy for a buyer's personal flood exposure history that will tend to be less reliable for floods in the more distant past relative to transaction date.

1.3.5 Disaster pressure index

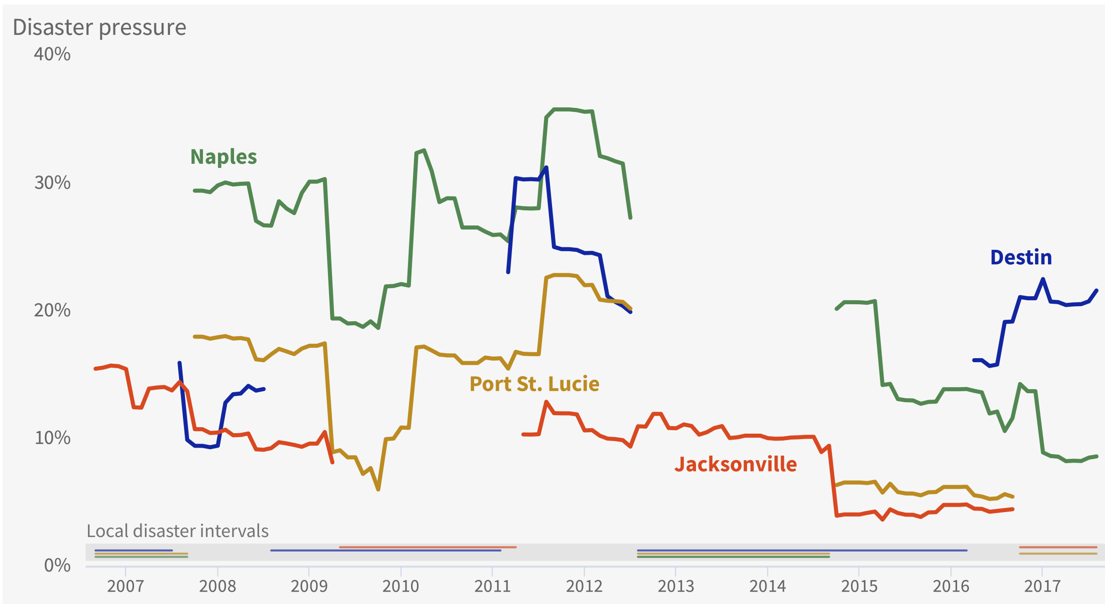
I use the same set of flood events to define a second treatment variable not at the individual transaction level but at the MSA-month level, which I call the "disaster pressure" index. This variable measures the share of buyers in a given MSA who have recently been exposed to flood events in their home counties. To calculate it, I consider the full group of buyers that appear in each MSA across my entire 13-year sample, whose combination of home counties form a catchment area for each MSA. I then tabulate the share of these counties, weighted by buyer population, that were treated in the last two years, as per the treatment definition above. By necessity, this exercise excludes same-address listers, so the denominator of the share calculation is the count of known-origin buyers. The disaster pressure index for MSA j in month t is given by:

$$Q_{t,j} = 100 \times \frac{\text{\# of known-origin buyers in MSA } j \text{ with local flood in the interval } (t-24, t)}{\text{\# of known-origin buyers in MSA } j}. \tag{1}$$

For example, a disaster pressure value of 10% for the Homosassa MSA in January 2010 indicates that 10% of the counties who generally send buyers to Homosassa experienced a flood disaster declaration in the 24 months leading up to January 2010, weighed by sender population. I intend for this index to capture the ambient level of disaster shock among the potential audience for real estate sales in a given area. In Appendix Table A1, I show

Figure 1.5 · The disaster pressure index in four MSAs over time

Disaster pressure over time in four representative MSAs from different regions of Florida. Disaster pressure is a measure of shock among the potential buyer population. It is defined in MSA *j* and month *t* as the share of the origin-determined buyer population for MSA *j* that experienced a flood shock in the 24 months preceding month *t*. I do not define disaster pressure in MSA *j* when that MSA has experienced a *local* flood shock in the previous two years; in these cases the disaster pressure variable would tend to be 70% or greater as the large majority of the buyer population in any MSA is comprised of local buyers. These periods are indicated by color-coded bars at the bottom of the figure.



that the results of my hedonic models are not sensitive to the twenty-four-month threshold in the definition of disaster pressure.

In every case, an MSA's catchment area is dominated by counties in the MSA itself, even when same-address buyers are excluded. Since local buyers make up 70-90% of buyers in most MSA-month markets, there is a natural ceiling on how many people can be shocked, absent a local flood event, and disaster pressure values tend to vary between 0% and 30% in my sample. This also implies that the disaster pressure index will spike whenever a local disaster strikes. I consider the disaster pressure index to be undefined in these cases and exclude them from my analysis, in order to better isolate the effect of remote disasters on local markets.

In Figure 1.5, I graph disaster pressure over time in four representative MSAs from different regions of Florida. Periods of local disaster when the index is undefined are indicated by color-coded bars at the bottom of the figure. While it is apparent that these measures are highly correlated across MSAs as counties that send many buyers to multiple MSAs are affected by disasters, each MSA has its own distinctive catchment area that leads to some variation in the disaster pressure index.

1.4 Retreat from the water: methodology and results

In this section, I lay out a hypothesis about how disaster shock will manifest in real estate buying behavior in Florida markets, propose a methodology to test this hypothesis, and show evidence in favor of the hypothesis from negative binomial count models.

I hypothesize that buyers who are disaster-shocked by recent floods near their homes will buy less property near the coast or in flood zones when they enter Florida coastal real estate markets than they would in the counterfactual non-shocked state. They will “retreat” from the coast, either opting for less flood-prone property, resorting to condo properties that partially decouple water access from flood risk, or exit the Florida coastal market entirely. One easily testable implication of this hypothesis is that disaster-shocked buyers will purchase homes that are farther from the water than otherwise similar buyers will. In this context, water proximity could be defined in terms of physical distance from the coastline, flood zone designation by the FEMA flood maps, or predicted sea-level-rise zones. Each is an imperfect proxy for true flood risk, and each may be used by disaster-shocked buyers as cues about where to avoid.

1.4.1 Model specification

To measure the causal effect of disaster shock in origin counties on Florida purchasing patterns, I estimate the retreat effect nonparametrically using a negative binomial regression in which the mean arrival rate of house purchases in a given area is given by:

$$\mu_{tdcj} = \exp(\alpha + \beta_{dl}(\mathbb{I}[d = D] \times \mathbb{I}[l = L]) + \gamma_l + \delta_d + \nu_c + \rho_j + \phi_t + \epsilon_{tdcj}), \tag{2}$$

where μ_{tdcj} is the number of properties sold at a distance d from the water, in month t and Florida MSA j, with buyers from non-Florida county c, β_{dl} are the coefficients of interest, and $\mathbb{I}[d = D]$ and $\mathbb{I}[l = L]$ are indicator variables for an observation of homes at distance interval d from the water, and purchased by people coming from counties that were last treated with a disaster l months ago, respectively. Also included are a constant term and fixed effects for time since disaster l, distance from the water d, county of origin c, Florida MSA j, month of sample t, and an error term ϵ_{tdcj}. A different measure of flood exposure can be substituted for d in Equation 2, such as sea-level-rise vulnerability or flood zone.

The unit of observation in the count models is the month-distance bin-buyer county of origin-Florida MSA. As an example, there are 13 transactions such that:

- the property is located in the Fort Myers MSA,
- the property sits less than 50 meters from the water,
- the buyers' county of origin is Cook County, Illinois,
- and the transaction occurred in April 2007.

These 13 transactions comprise one observation. I elect to use a negative binomial model as opposed to a Poisson model because approximately 97% of these observation bins have zero home purchases, indicating significant overdispersion.

My identifying assumption is that, conditional on month and county, the occurrence of disaster declarations is random with respect to home purchasing behavior. That is, factors that would affect home purchasing behavior do not vary systematically with disaster timing within a county, controlling for time of year. Under this assumption, the γ_l and β_{dl} coefficients represent the causal effect of a disaster declaration in a given county on Florida home purchases in distance interval d by residents of that county l months later, relative to purchases in the excluded distance category (5+ km). While this staggered difference-in-difference, two-way fixed effects (TWFE) design is susceptible to the bias identified by Baker, Larcker and Wang (2022) and Goodman-Bacon (2021), the fact that the treatment is not permanently applied after one disaster, but is instead applied temporarily after each flood and removed thereafter, should attenuate this concern.

The total causal effect identified may comprise wealth effects from disaster damage, belief shocks, and preference shocks (intensive margin), and selection out of migration to Florida (extensive margin). Therefore, my measure of the effects of interest, changes to risk preference and beliefs, are at risk of contamination. In order to reduce this contamination, I conduct robustness checks wherein I exclude from the analysis any ZIP codes where the count of flood insurance claims in the month of a disaster declaration is abnormally high; see Appendix A3 for a description of how I determine this. Figure A2 illustrates this procedure with a hypothetical disaster.

However, this approach does not eliminate the possibility of a selection effect driving my results. In principle, a positive finding of retreat could be driven by a compositional change in the buyers entering the Florida market, as opposed to a change in any individual agent's beliefs or preference over time. This would suggest a more complicated mechanism at work, where a flood event in County *X* differentially motivates certain residents with different preferences to buy homes in Florida, as opposed to the availability bias/Bayesian updating explanation above. With this identification strategy, I am unable to distinguish between these two explanations. However, they have similar implications for affected markets in Florida. Additionally, the potential for the most shocked buyers to differentially exit the market biases my estimates of any effect downward, as the remaining shocked buyers who enter the market are presumably less scarred than the average county resident.

1.4.2 Results from count models

Table 1.4 displays negative binomial regression estimates showing evidence of retreat from the water in three different dimensions: straight-line distance, flood zone, and sea-level-rise zone. Each panel reports results from a single regression; the columns report results from different groups of interaction terms. Each model was estimated using a full set of *treatment* × *distance* interactions and main effects, as well as fixed effects at the county and month-of-sample levels. The models in this table are estimated on the sample described in Table 1.1, Column 4. Note that for the model shown in Panel A, all condominium properties are included in the "condo" distance category regardless of their physical distance from the water.

In this main sample, I find strong evidence of retreat from low-lying areas in Florida real estate markets by disaster-shocked buyers on the intensive margin. The results in the top panel suggest that buyers from shocked counties decrease their non-condo home purchasing in the areas near the water by 20%-30%, relative to their purchases of properties farther inland, in the months after a disaster. I also find that condo sales are relatively depressed at first, but less so than non-condo properties near the water, and eventually (or more precisely, among buyers who have gone longer since their treatment shock) condo sales are higher relative to inland properties. I also find a rebound effect as most coefficients turn positive for buyers who were shocked more than 3 years ago, but less than 5. The models displayed in Panel B and Panel C, where I substitute flood zone and sea-level-rise zone bins for distance bins, tell a similar story. Figures 1.6 and 1.7 display the coefficients from Panel A and Panel B graphically and illustrate the strong gradient in homebuying behavior introduced by disasters in buyers' home counties.

One major finding goes against my hypothesis: I find scattered evidence of a rebound effect where shocked buyers begin buying *more* flood-prone properties 2-5 years after the shock, relative to non-shocked buyers. This could reflect pent-up demand by buyers hailing from shocked counties who waited until the shock receded before plunging back in to the market. Overall, the shift in coefficients across the columns of Table 1.4 tracks with several other findings in the literature made in the context of flood insurance (Gallagher 2014; Kousky 2017), earthquake insurance (Palmquist 1992), and climate concern (Konisky, Hughes and Kaylor 2016). A major contribution of the present work is documenting this pattern – and its troubling implications for the ability of personal adaptation to ameliorate the effects of climate change – in a new and important domain. Figure 1.8 illustrates this using the results from the first row of Panel C, and shows a contrast between the full sample and a subsample of out-of-state buyers, discussed more below.

This result may also partially reflect the fact that treatment is measured with increasing error for storms in the more distant past, due to the fact that I cannot observe buyers' counties of residence at any time before the time of their home purchases. For this reason, results for the treatment groups on the right-hand side of the table should be interpreted with more caution.

Table 1.4 · Sales count models

	Treatment group (months since flood disaster)					
	0-6	*6-12*	*12-24*	*24-36*	*36-48*	*48-60*
Panel A. Distance from water						
Frontage (<50 m)	-0.208***	-0.273***	-0.0814***	0.0190	0.0818***	0.0560**
	(0.0197)	(0.0189)	(0.0177)	(0.0202)	(0.0228)	(0.0238)
Condo	-0.0833***	-0.115***	0.0881***	0.132***	0.0995***	0.0928***
	(0.0115)	(0.0109)	(0.0103)	(0.0118)	(0.0137)	(0.0144)
100 m	-0.189***	-0.291***	-0.0940***	0.0560**	0.0938***	0.0722**
	(0.0249)	(0.0241)	(0.0228)	(0.0255)	(0.0293)	(0.0313)
200 m	-0.158***	-0.255***	-0.0623***	0.0432**	0.0761***	0.0758***
	(0.0211)	(0.0202)	(0.0192)	(0.0216)	(0.0246)	(0.0262)
300 m	-0.185***	-0.257***	-0.0790***	0.0579**	0.0883***	0.0862***
	(0.0231)	(0.0225)	(0.0214)	(0.0241)	(0.0276)	(0.0290)
400 m	-0.205***	-0.305***	-0.0759***	0.0802***	0.0932***	0.0846***
	(0.0254)	(0.0247)	(0.0235)	(0.0264)	(0.0304)	(0.0322)
500 m	-0.183***	-0.298***	-0.0651**	0.0696**	0.127***	0.107***
	(0.0277)	(0.0268)	(0.0254)	(0.0289)	(0.0326)	(0.0347)
600 m	-0.204***	-0.322***	-0.0817***	0.0552*	0.114***	0.125***
	(0.0292)	(0.0286)	(0.0273)	(0.0308)	(0.0349)	(0.0371)
700 m	-0.212***	-0.318***	-0.0926***	0.0580*	0.0819**	0.124***
	(0.0306)	(0.0300)	(0.0288)	(0.0324)	(0.0372)	(0.0396)
800 m	-0.244***	-0.352***	-0.106***	0.0800**	0.0709*	0.119***
	(0.0327)	(0.0322)	(0.0303)	(0.0345)	(0.0398)	(0.0412)
900 m	-0.242***	-0.352***	-0.0685**	0.0620*	0.111***	0.126***
	(0.0340)	(0.0329)	(0.0311)	(0.0358)	(0.0402)	(0.0431)
1000 m	-0.284***	-0.356***	-0.0915***	0.0472	0.0808**	0.0763*
	(0.0348)	(0.0337)	(0.0319)	(0.0362)	(0.0411)	(0.0440)
Omitted category: 5-10 km (some categories omitted for brevity). Observations: 963,412 transactions, 615,302 count cells. Wald $\chi^2(155) = 66578$.						
Panel B. Flood zones						
VE (100-yr flood + storm surge)	-0.175***	-0.220***	-0.108***	0.0157	0.0794**	0.00134
	(0.0296)	(0.0283)	(0.0257)	(0.0283)	(0.0317)	(0.0345)
A (100-yr flood plain)	-0.0945***	-0.0918***	-0.0371***	0.00349	0.0269**	0.0367***
	(0.0104)	(0.00999)	(0.00910)	(0.0102)	(0.0116)	(0.0122)
AH (100-yr shallow flooding)	-0.201***	-0.258***	-0.135***	-0.0522	-0.0172	0.0473
	(0.0360)	(0.0347)	(0.0311)	(0.0347)	(0.0391)	(0.0427)
X (500-yr flood plain)	-0.112***	-0.156***	-0.0579***	0.0433***	0.0711***	0.0724***
	(0.0145)	(0.0141)	(0.0128)	(0.0142)	(0.0163)	(0.0172)
Omitted category: outside 500-yr zone. Observations: 963,412 transactions, 468,233 count cells. Wald $\chi^2(71) = 56984$.						
Panel C. Projected sea level rise						
<6-ft SLR	-0.132***	-0.171***	-0.0677***	0.0297**	0.0620***	0.0409**
	(0.0141)	(0.0134)	(0.0123)	(0.0136)	(0.0154)	(0.0163)
6-ft SLR	-0.131***	-0.160***	-0.0647***	0.00799	0.0666***	0.0365*
	(0.0162)	(0.0154)	(0.0140)	(0.0158)	(0.0179)	(0.0188)
Omitted category: outside SLR zone. Observations: 963,412 transactions, 422,280 count cells. Wald $\chi^2(57) = 66748$.						

Negative binomial estimates with robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is counts of sales at the month-proximity bin-buyer county of origin-Florida MSA level. Each panel reports results from a single regression; the columns report results from different groups of interaction terms. The number of cells varies across regressions as the bins are defined according to a different flood risk factor in each case (distance in Panel A, flood zone in Panel B, and sea-level-rise submersion zone in Panel C). For each observation, the treatment category is defined by the most recent flood event in the buyer's county of origin at the time of sale. Each model was estimated using a full set of *treatment* × *distance* interactions, as well as fixed effects at the county and month-of-sample levels. For the model shown in Panel A, condominium properties are included in the "condo" distance category regardless of physical distance from the water.

Table 1.5 · Sales count model robustness checks

	(1) Full sample	(2) No flooded ZIP codes	(3) No data issues	(4) No condos	(5) No owner-occupiers	(6) No small counties	(7) No intra-FL buyers
6-12 month treatment group							
VE (100-yr flood + storm surge)	-0.220***	-0.195***	-0.155***	-0.213***	-0.208***	-0.236***	-0.223***
	(0.0283)	(0.0295)	(0.0323)	(0.0541)	(0.0326)	(0.0312)	(0.0305)
A (100-yr flood plain)	-0.0918***	-0.0824***	-0.0678***	-0.0944***	-0.0715***	-0.0840***	-0.0834***
	(0.00999)	(0.0103)	(0.0109)	(0.0131)	(0.0120)	(0.0110)	(0.0105)
AH (100-yr shallow flooding)	-0.258***	-0.250***	-0.204***	-0.277***	-0.287***	-0.258***	-0.238***
	(0.0347)	(0.0366)	(0.0357)	(0.0481)	(0.0460)	(0.0363)	(0.0377)
X (500-yr flood plain)	-0.156***	-0.150***	-0.111***	-0.199***	-0.164***	-0.145***	-0.155***
	(0.0141)	(0.0147)	(0.0151)	(0.0185)	(0.0179)	(0.0153)	(0.0152)
48-60 month treatment group							
VE (100-yr flood + storm surge)	0.00134	0.0806**	0.0461	0.0958	0.0393	0.00625	-0.115***
	(0.0345)	(0.0363)	(0.0396)	(0.0667)	(0.0401)	(0.0381)	(0.0363)
A (100-yr flood plain)	0.0367***	0.0571***	0.0581***	0.0554***	0.0628***	0.0561***	-0.0330***
	(0.0122)	(0.0127)	(0.0135)	(0.0163)	(0.0149)	(0.0136)	(0.0125)
AH (100-yr shallow flooding)	0.0473	0.0697	0.0662	0.124**	0.101*	0.0631	-0.0696
	(0.0427)	(0.0454)	(0.0448)	(0.0577)	(0.0552)	(0.0448)	(0.0450)
X (500-yr flood plain)	0.0724***	0.107***	0.0955***	0.103***	0.101***	0.0951***	-0.0249
	(0.0172)	(0.0181)	(0.0186)	(0.0225)	(0.0220)	(0.0188)	(0.0181)
Transactions	963,412	873,551	785,751	627,174	541,883	852,307	710,931
Share of full sample	100%	91%	82%	65%	56%	88%	74%
Count cells	468,502	437,942	412,619	319,343	307,839	369,494	404,114
Origin counties	818	818	818	815	818	306	760
Wald test: $\chi^2(71)$	56985	50449	48575	32707	23317	57176	58538

Negative binomial estimates with robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is counts of sales at the month-proximity bin-buyer county of origin-Florida MSA level. Each column reports results from a regression with a portion of the sample excluded. Column 2: see Appendix A2 for information on how flooded ZIP codes were identified. Column 3: observations with data anomalies were excluded (see Appendix A1). Column 6: buyers from counties with fewer than 1,000 total buyers across the sample are excluded. Each model was estimated using a full set of *treatment* × *distance* interactions, as well as fixed effects at the county and month-of-sample levels, but only the interactions for the 6-12 month and the 48-60 month treatment groups are displayed. The results displayed in the Column 1 are from the same model reported in Panel B of Table 1.4.

Figure 1.6 · Retreat effect among 6-12 month treatment group: distance from water

Point estimates and 95% confidence intervals from the model displayed in Table 1.4, Panel A. Buyers shocked in the past six months purchase significantly fewer homes in the 50-4,000m range, relative to counterfactual buyers from the same county who are not recently shocked.

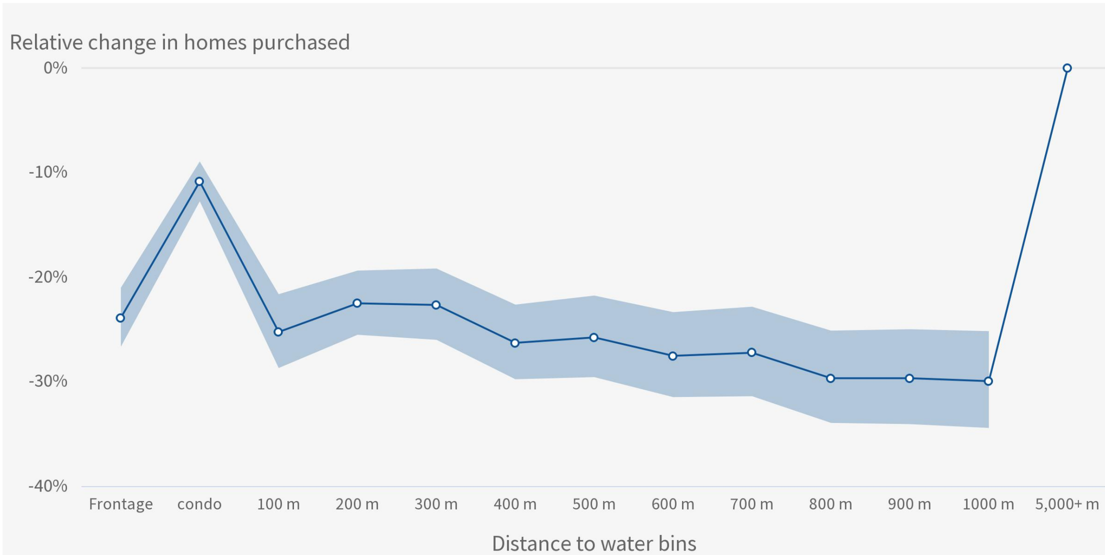


Figure 1.7 · Retreat effect among 6-12 month treatment group: flood zones

Point estimates and 95% confidence intervals from the model displayed in Table 1.4, Panel B. Buyers shocked in the past six months purchase significantly fewer homes in flood zones, relative to counterfactual buyers from the same county who are not recently shocked.

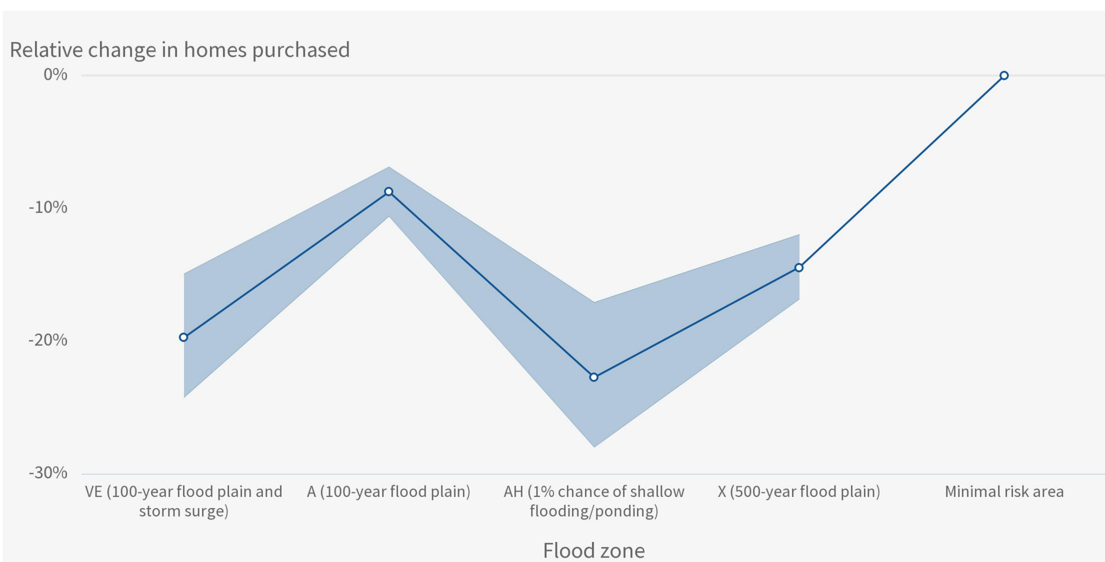
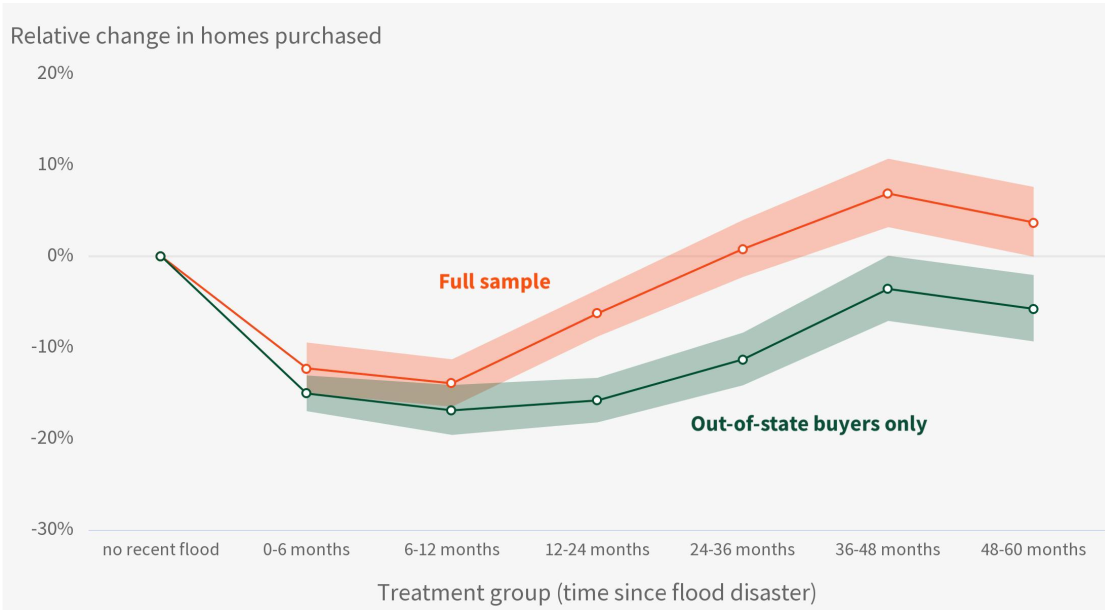


Figure 1.8 · Transience of the retreat effect in the 6-ft sea-level-rise zone

Point estimates of the interaction with 6-ft sea level rise zone for treatment groups with differential recency of exposure, relative to the untreated group with no disasters in the past five years. The effect diminishes over time as each successive group shows less and less retreat behavior, and even reverses in the full sample. Shaded areas display 95% confidence intervals. The estimates in orange match the first row of Table 1.4, Panel C.



Overall, this response could be explained most simply by the availability heuristic: the tendency to perceive events as more likely when they can be more easily imagined, remembered, or brought to mind (Slovic, Kunreuther and White 1974; Tversky and Kahneman 1973, 1974). This is the explanation favored by Dessaint and Matray (2017) in the context of managers' response to hurricanes. Flooding risk will loom large in the minds of shocked buyers as they consider their options in Florida real estate markets, they will feel more trepidation about living near the water than they would have absent the shock, and they will be more willing to spend on homes that feature other amenities beside water proximity.

Alternatively, a simple Bayesian updating model may be able to explain retreat from the water, although given the relative frequency of flood events it would require significant forgetting for one storm to change behavior in a measurable way (Gallagher 2014), and would require transference of learning from the home county to the Florida context, as well as an assumption that floods in distant counties are providing informative updates about flood risk in Florida.

Assuming water proximity is a valuable amenity in this context, a negative wealth effect could also explain retreat behavior, even absent any changes in beliefs about floods or availability bias. I take steps to remove buyers who may have been directly hit by flood events and experienced negative wealth shocks (see Appendix A2), as well as other groups of observations that may be driving a spurious result.

Table 1.5 displays abridged results from six additional negative binomial models, each of which is run on a subsample with certain transactions excluded. Each model was estimated using a full set of *treatment* × *distance* interactions, as well as fixed effects at the county and month-of-sample levels, but only the interactions for the 6-12 month and the 48-60 month treatment groups are displayed. Note that the results displayed in the Column 1 are from the same model reported in Panel B of Table 1.4.

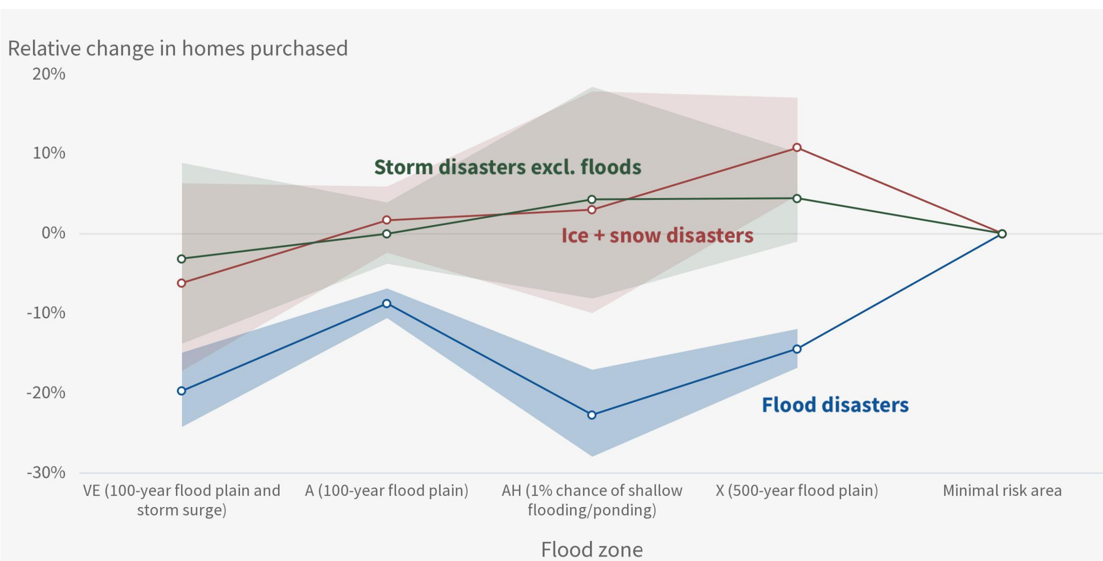
Table 1.6 · Falsification tests for count models

	(1)	(2)	(3)
		Disaster category	
	Flood	Snow/ice	Severe storm
48-60 month treatment group			
VE (100-yr flood + storm surge)	-0.220***	-0.064	-0.0324
	(0.0283)	-0.0638	-0.0596
A (100-yr flood plain)	-0.0918***	0.0164	-0.000323
	(0.00999)	-0.0208	-0.0196
AH (100-yr shallow flooding)	-0.258***	0.0294	0.0419
	(0.0347)	-0.0687	-0.0648
X (500-yr flood plain)	-0.156***	0.102***	0.0433
	(0.0141)	-0.0282	-0.0273
Transactions	963,412	963,412	963,412
Count cells	472,736	472,736	472,736
Origin counties	818	818	818
Wald test: $\chi^2(71)$	57029	57910	56425

Negative binomial estimates with robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is counts of sales at the month-proximity bin-buyer county of origin-Florida MSA level. Each column reports results from a regression with a different treatment definition. Each model was estimated using a full set of *treatment* × *distance* interactions, as well as fixed effects at the county and month-of-sample levels, but only the interactions for the 6-12 month and displayed. The results displayed in the Column 1 are from the same model reported in Panel B of Table 1.4.

Figure 1.9 · Retreat from the water under different disaster definitions

Point estimates and 95% confidence intervals from the models displayed in Table 1.6. See text in Section 1.4.3 for more on how each disaster category is defined.



1.4.3 Falsification tests

I perform two falsification tests – tests where my expect a null result given my hypothesis – to bolster my analysis. In each test, I estimate my main specification from Equation 2 using a placebo version of my treatment variable.

In the first test, I use snow and ice disasters in place of flooding disasters. Specifically, I include all disasters classified by FEMA as “severe ice storm,” “snow,” and “freezing.” Additionally, I include disasters whose official descriptions include the terms “ice,” “snow,” or “winter.” Finally, I exclude a small number of disasters that satisfy these criteria but also count as flood disasters according to my definition (see Table 1.3), leaving this set of disasters disjoint from the set of flooding disasters in the main analysis. I then reassign the treatment variable for each transaction based on the time since the most recent snow or ice disaster, rather than the most recent flooding disaster.

In the second test, I use a similar approach, replacing flooding disasters with all storms classified by FEMA as a “severe storm,” excluding disasters that count as flood disasters. This set of disasters includes storms with heavy rain, straight-line (i.e. not hurricane) winds, and tornadoes.

According to my hypothesis, exposure to these storms should not affect demand for waterfront property in Florida relative to properties farther inland, although plausibly could affect demand for real estate in warm climates in the case of ice storms. In terms of Equation 2, I would not expect the estimates for β_{dl} to be statistically significant. Table 1.6 displays results from estimates for the two falsification tests alongside results from the main model (Column 1). With one exception, none of the coefficients are statistically significantly different from zero, and they are also generally quite significantly different from the results in Column 1. This provides reassurance that the effect found is not an artifact and is driven by flood experience specifically. Figure 1.9 illustrates these same estimates graphically.

1.5 Sorting model and welfare analysis

The reduced-form evidence from the previous section tells an important story about retreat, but cannot put a valuation on the size of the retreat effect. In this section, I employ a sorting model that I use to estimate the difference in marginal willingness to pay (MWTP) for water access between shocked and non-shocked buyers. Below, I introduce the sorting model and my matching strategy that I argue allows me to interpret differences across treatment groups as causally related to my treatment variable.

1.5.1 Model construction

Each household has unit demand for housing in a particular Florida housing market, and chooses from among J different types of housing where homes in each type are nearly identical on observables, following Bakkensen and Ma (2020). I call each housing group j a “residence.” For any residence in the household’s consideration set, the household decides whether it would purchase a full insurance contract with annual premium I_j conditional on purchase. Thus, the indirect utility function for household i selecting residence j is:

$$V_j^i = \alpha_x^i X_j - \beta(P_j^r + \min[I_j, \gamma f_j^i (1/r) C_j^i]) + \xi_j + \epsilon_j^i, \tag{3}$$

where:

- X_j is a vector of observed characteristics about each residence, including distance to the water
- P_j^r is the annuitized value of the home at interest rate r
- I_j is the premium charged by NFIP for the residence,
- f_j^i is the annual likelihood of a flood for residence j, as perceived by household i
- $(1/r)C_j^i$ is the expected wealth loss from a flood event as perceived by the buyer,

- γ is an adjustment to account for risk preference, such that $-\gamma f_j^i(1/r)C_j^i$ is the certainty equivalent of the lottery induced by flood risk,

- ξ represents unobserved characteristics of each residence, and

- ϵ_j^i represents idiosyncratic preferences.

The household selects the residence that maximizes its indirect utility, and this choice (although not the insurance purchase decision) is observed by the researcher. Note that α_x^i is indexed at the household level, admitting the possibility of different preferences for different individuals (or groups). The utility cost of annual payments β is constant across all households.

The model presented above incorporates several assumptions. First is that demand for housing in a given time period and housing market is exogenous; the decision process by which households decide to relocate to Florida in a particular time period is determined by factors outside the model, most crucially experienced flooding. This assumption would almost certainly not hold in a few cases where people suffered property damage from floods and incurred a huge loss in wealth, or had an imperative to find a new home.

I estimate a model that is simplified in many dimensions. First, I proxy flood exposure by distance to the water, and ignore the considerations of insurance costs varying across areas. I also use a limited set of features to define the residences (building size quartiles, flood zone, pool, condo indicator, and census tract), so each residence may still be relatively heterogeneous. Finally, I acknowledge that the assumption of fixed demand by households is not realistic, as many households may be selecting in or out of the Florida market based on flood disaster history or any of innumerable other factors. Nonetheless, I argue that the population of buyers that experience a flood and nonetheless enter the sample are less deterred by flood disaster exposure, which biases my estimates of any differences toward zero.

I follow Bakkensen and Ma (2020) in allowing for heterogeneous tastes for some variables in X_j across disaster shock status. I group households by the recency of flood disasters in their counties of origin. One group is comprised of households who were shocked in the last 24 months before purchase, another group those who were shocked in the last 36 months but not the last 24 months, etc. The omitted group is comprised of households with no recent shock, specifically no flood in the last 5 years. This creates a partially random partition of $K = 6$ groups who should have systematically different preferences according to my various hypotheses.

I assume that tastes are homogeneous across groups, with the key exception of proximity to the water. Assume D_j in some measure of distance from the water, whether literal straight-line distance or defined in terms of flood zones or sea-level-rise zones. Again, this elides some of the complexity from Equation 3. I rewrite the indirect utility function for household i in group z from neighborhood j as follows:

$$V_j^{z\in Z} = \delta_j + \alpha^z D_j + \epsilon_j^i, \tag{4}$$

where the δ_j term represents the mean utility of the base group (those without recent disasters) associated with residence j, inclusive of rent costs, and is a function of other residence characteristics X_j such as square footage, and structure type, as well as distance from the water, the average annualized rent p_j, and a tract-level fixed effect ξ_c:

$$\delta_j = \beta D_j - \alpha_p p_j + \gamma_x X_j + \xi_c + \varepsilon_j. \tag{5}$$

The coefficient α^k is a measure of the additional marginal utility associated with an increase in D_j for group k, relative to the base group. Meanwhile, β is an estimate of the marginal utility of that base group. As proximity to the coast is generally considered to be desirable, I would expect $\beta < 0$. However, disaster-shocked groups are likely to find the coast relatively unappealing, so I would also expect $\alpha^z > 0$ for all $z \in Z$ outside the base group.

I proceed to estimate these parameters in two stages. First, I estimate a^z for each group with an expectation-maximization algorithm. Based on the observed choice data, I use the contracting mapping proposed by Berry (1994) to find a set of mean utilities δ_j that rationalize the choice probabilities across the sample (expectation step). Then I estimate the a^z values with an alternative-specific conditional logit model, maximizing the following log-likelihood given δ_j's (maximization step):

$$L = \sum_i^N \sum_j^J \mathbb{I}(d^i = j) \log P_j^i$$

where P_j^i is the probability that household i chooses residence j, conditional on the δ_j's, α^z's, and the taste shock ϵ_j^i having a Type I Extreme Value distribution:

$$P_j^i = \mathbb{P}[V_j^i \geq V_{j'}^i \forall j' \neq j | \delta, \alpha] = \frac{e^{V_j^i}}{\sum_{j'} e^{V_{j'}^i}}. \tag{6}$$

One advantage of this approach is I can estimate α^z using the EM procedure without first estimating Equation 5. This yields computational savings and also allows me to cabin the identification challenges inherent in estimating that equation, where price is sure to be correlated with the unobserved error ε_j.

1.5.2 Instrumental variables approach

A precise estimate of α_j is vital to translate the model from utility terms to dollars. To address the endogeneity in Equation 5 and obtain consistent estimates of β, α_p and γ_x, I subsequently instrument for price using two strategies. First, I directly instrument for residence sales price in census tract c with property characteristics of all census tracts whose centroids are within 5km of the centroid of tract c, including factors like average square footage and building age in those nearby tracts (hereinafter Z_c). The exclusion restriction is satisfied as long as the characteristics of those nearby communities do not directly affect the utility that one derives from living in census tract c. Note that X_j does include measures of local amenities like recreational areas that stretch beyond the borders of the residence itself, so the model accommodates the possibility that nearby amenities will affect valuations across tract boundaries. However, the IV strategy assumes that the characteristics of the houses themselves do not affect neighboring valuations in this way.

Second, I adopt the procedure proposed by Bayer and Timmins (2007) to construct a univariate instrument for price. Let $\hat{\delta}$ indicate the estimates of each neighborhood mean utility obtained from the first stage estimation. To do so, I first estimate a naive OLS model to obtain an (inconsistent) estimate for α_p. This will serve as my initial guess of the price coefficient in an iterative procedure. I then create price-adjusted δ values which I denote with a tilde:

$$\tilde{\delta}^{(0)_j} = \hat{\delta}_{j_j} + \hat{\alpha_p}^{(0)} p_j,$$

where the superscript (0) denotes the zeroth iteration of the procedure. I then estimate a model of this price-adjusted $\tilde{\delta}$ with the same set of controls X_j along with the additional controls capturing the characteristics of surrounding census tracts Z_c:

$$\tilde{\delta}_j = \beta D_j + \gamma_x X_j + \gamma_z Z_c + \xi_c + \varepsilon_j. \tag{7}$$

I estimate Equation 7 by OLS to obtain a fitted value for $\tilde{\delta}_j$, which I call $\bar{\tilde{\delta}}$. Finally, the instrument is obtained by finding the price vector that rationalizes the $\bar{\tilde{\delta}}$ given $\hat{\alpha_p}^{(0)}$. This is calculated by:

$$p_{j,\text{iv}}^{(0)} = p_j + \frac{\bar{\tilde{\delta}}_j - \tilde{\delta}_j}{\hat{\alpha_p}^{(0)}} \tag{8}$$

In essence, the instrument captures the variation in $\bar{\tilde{\delta}}$ not reflected in $\tilde{\delta}$, which in turn reflects information about the correlation of the error term ε_j with the neighboring tract data Z_c. I then instrument for price using p_{iv} and obtain an estimate for α_p which serves as the new value $\hat{\alpha_p}^{(1)}$ in the next round of iteration. I continue until

$\hat{\alpha_p}$ converges to within tolerance.

1.5.3 Genetic matching procedure

Estimated on my partition of Z groups defined by local disaster recency, the model returns estimates of α^z indicating the differential valuation of water proximity among each group. But while I treat flooding arrivals as a random process, it is not distributed uniformly across the country. This means that people from certain flood-prone counties are mechanically more likely to end up in the treated group. The people living in these counties may be different in all sorts of ways, and in particular in their preference for living near water.

This means that I cannot interpret the α^z's causally, because any disaster shock effect will be comingled with clear underlying preference differences between the groups. I would expect this to bias the magnitude of the α^z's downward relative to the true causal value, as people living in flood-prone areas have already demonstrated a relative preference for water proximity.

In order to isolate α^z that reflect behavior differences causally related to disaster shock, I employ a genetic matching procedure. Genetic matching is a computationally-intensive evolutionary optimization procedure that uses successive "generations" of proposed weighting matrices to minimize the Mahalanobis distance between the treated and untreated samples (Diamond and Sekhon 2013). Genetic matching was performed using the MatchIt package in R (Stuart et al. 2011), which calls functions from the Matching package (Diamond and Sekhon 2013; Mebane Jr and Sekhon 2011). The aim of this procedure is to find subsamples of the treated and untreated groups that are similar on a key set of variables, both in terms of means and higher moments.

I focus on only two of the Z groups defined in Section 1.5.1: buyers shocked within the last 24 months (treated) and buyers with no flood shocks in the past five years (untreated). Judging by the results of the count models in Section 1.4, this is the group whose behavior is most profoundly impacted by flood exposure (see Table 1.4). Matching across multiple treated groups greatly complicates the matching process and can whittle effective sample sizes down to nothing as more and more observations are set aside. This effectively excludes people in the 24-60 month treatment groups, but allows for the sharpest comparison and the most credibly matched sample possible.

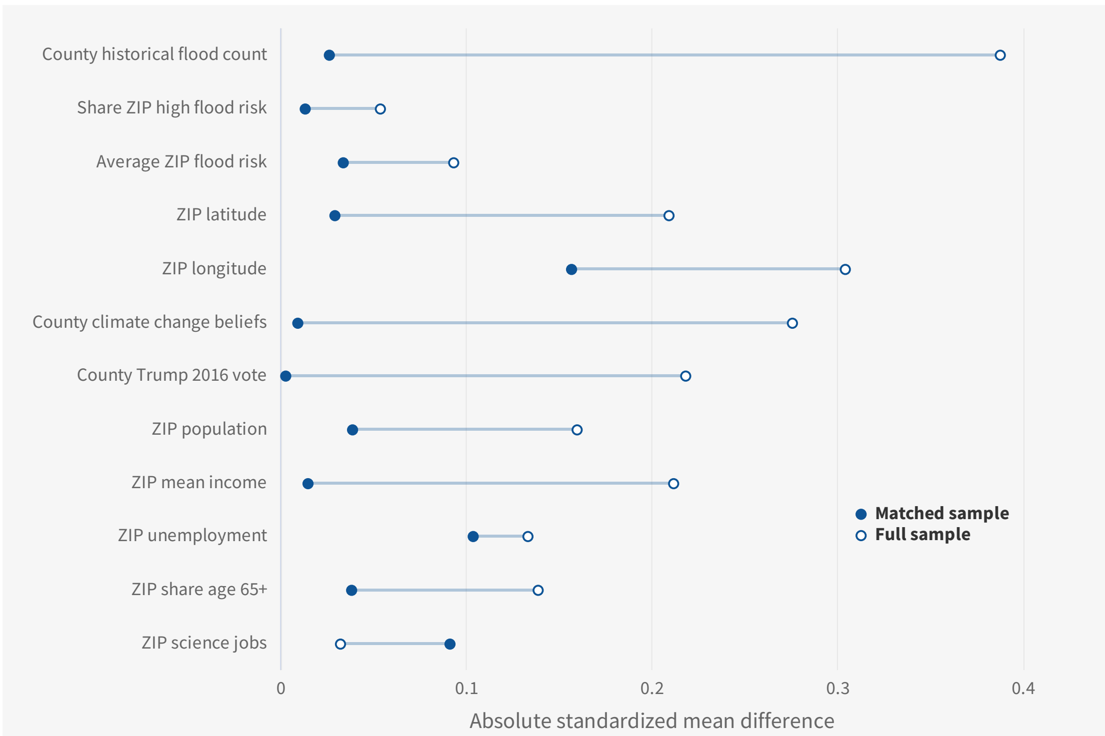
I use twelve variables as the basis for matching, including three key variables pertaining to local flood risk: the historical number of flood disasters in the county since 1960, and two measures of flood risk at the ZIP code level published by First Street Foundation (2020). These variables vary only on the ZIP code (and sometimes county) level, so the matching procedure effectively isolates pairs of ZIP codes that are similar on these observable traits, but received different levels of treatment. Each MSA subsample is matched separately, creating 15 matched sets of observations with similar characteristics. A Love plot of absolute standardized mean differences (ASMD) for these variables before and after matching is displayed in Figure 1.10. To calculate the ASMD, I divide the absolute difference in means for each variable by the standard deviation of the untreated group in the full, unmatched sample. I use this same pre-match standard deviation quantity to calculate the ASMD of the matched sample. Appendix Figure A3 contains QQ plots that show the algorithm successfully matches the distribution of each variable, not just the means.

As expected, the unmatched samples vary greatly in terms of historical flood count, as that is the basis for the treatment variable in the first place. The ZIP code flood measures from First Street Foundation are more balanced to begin with; this likely reflects the fact that a treated county will contain a mix of relatively flood-prone and flood-immune areas, all of which are included together in the treatment group.

The matching procedure successfully isolates a matched sample where these differences are much less pronounced. The only variables on which the match fails to secure a tight match are the ZIP latitude and ZIP longitude variables, which reflect the fact that flood risk is concentrated in certain parts of the country. The extent to which ZIP latitude and longitude failed to converge varies across different MSA samples, but the Tampa MSA data in Figure 1.10 is fairly representative in this regard. Nonetheless, the excellent balance on flood risk variables indicates the matched sample is composed of groups facing quite similar levels of flood risk; their assignment to the treatment group conditional on this restriction is close to random.

Figure 1.10 · Results of genetic matching on data from Tampa MSA

The Love plot below indicates the absolute standardized mean differences in the twelve variables used to match in the genetic matching algorithm. To calculate the absolute standardized mean difference (ASMD), I divide the absolute difference in means for each variable by the standard deviation of the untreated group in the full, unmatched sample. I use this same pre-match standard deviation quantity to calculate the ASMD of the matched sample. Appendix Figure A3 contains QQ plots that show the algorithm successfully matches the distribution of each variable, not just the means. Data sources: U.S. Census American Community Survey, FEMA disaster records, First Street Foundation (ZIP flood risk), MIT Election Data and Science Lab (Trump 2016 vote), Yale Program on Climate Change Communication (county climate change beliefs).



1.5.4 Model results

To reduce computational burden, I narrow my focus to five of the MSAs with the greatest potential policymaking interest: the three largest MSAs in Florida (Miami, Tampa, and Jacksonville), along with the MSA with the greatest share of out-of-state buyers (Naples) and the largest MSA on the Florida panhandle, which is otherwise unrepresented (Destin).

Table 1.7 displays these results on two sets of samples: a full sample and a matched sample for each MSA. Each row displays results from a single model. Columns 1 and 2 are presented in utility units, and are not interpretable outside the model. The estimates for α_p (Column 3) show the marginal utility of a thousand dollars in annual rent, also in utility terms, while the figures in Columns 4 and 5 are calculated using this marginal utility figure. Standard errors are bootstrapped for all estimates using 50 replications.

In each case, I use both IV strategies discussed above in attempts to get the most precise and reasonable estimate of α_p possible. In many cases, one or the other IV strategy yields an estimate of α_p that is positive, which implies a misspecified model or lurking endogeneity. Although there is some variation across the bootstrapped samples, the direct instruments generally passed the weak instruments test and the Sargan-Hansen test of overidentifying restrictions (Sargan 1958) handily, but in many cases not the Durbin-Wu-Hausman test, calling the consistency of these estimates into question (Hausman 1978).

Aside from these concerns, the estimates are very noisy and in some cases, highly implausible. It is possible that additional bootstrap replications or more fine-grained residence definitions would improve the signal-to-noise ratio. Broadly speaking, the results suggest that the differential valuation of flood zone status by shocked buyers is in the range of \$500 to \$3,000 per year. Given the average home sales price in the sample, this is in the range of 1% of home value per year.

1.5.5 Welfare analysis

The analysis from the sorting models suggests that shocked homebuyers sort differently from their non-shocked counterfactual selves, attributing less value to flood-prone residences. Is this adaptive behavior significantly improving consumers' choices on their own terms? To answer this question, I must take a stand on whether the choices made after a disaster shock, or instead the choices made with no recent disaster weighing on the mind, represent welfare-relevant preferences.

First, I assume the new amenity values are mistaken – the result of post-disaster shock or hysteria that will recede over time as individuals come to their senses. This accords with the framework of Bernheim and Rangel (2009), in which choices made within "suspect" choice environments should not contribute to welfare analysis. In that framework, recent flood exposure at home is an ancillary condition that may influence choice but is not relevant to a hypothetical central planner. They define a mistake as a choice made in a suspect choice environment that is contradicted by choices in non-suspect environments. In this sense, the retreat from the water could be interpreted as a mistake, because the shocked buyers are not making the same choices as their counterfactual selves. In that case, the flood disaster treatment distorts agents' valuations of each neighborhood v_j, and leaves the average homebuyer worse off when the effect subsides and they realize the true v_j of each potential home they could have bought.

Hanemann (1982) finds that the compensating variation of a change in v_j in this context can be approximated by:

$$\text{cv} = \frac{1}{\gamma}\left[\log\sum_j e^{v_j^{\text{new}}} - \log\sum_j e^{v_j^{\text{old}}}\right] \tag{9}$$

This approach is often employed when analyzing changes to the value of residential choices induced by a pollution release or an improved environmental amenity. My setting is different: the change in v_j induced by the flood treatment does not represent a change in the local environment – nothing about the neighborhoods in Florida is changing as the result of a distant flood, and they provide buyers with no more or less value than they did before. However, if some or all of the agents are optimizing according to new v_j, in some sense this is analogous to a change in local circumstances. Leggett (2002) proposes this modification where correct valuations are given by v_j^c and mistaken valuations are given by v_j^m:

$$\text{cv} = \frac{1}{\gamma}\left[\left(\log\sum_j e^{v_j^{\text{m}}} - \sum_j \pi_j^m(v_j^m - v_j^c)\right) - \log\sum_j e^{v_j^{\text{c}}}\right]. \tag{10}$$

The term highlighted in red captures both the mis-optimization that occurs when the shocked agents maximize with distorted v_j's, and also the displacement of other people from preferable options that are chosen by the mistaken optimizers. To implement this analysis, I take the estimated utility values of residences as v_c directly from the model and add a^z to those valuations to obtain v_m: the mistaken valuations of residences by the shocked population. The sorting models displayed in Table 1.7 provide estimates of both v_c and v_m that allow me to calculate the approximate compensating variation of the flood disaster treatment.

Alternatively, I could assume that it is the shocked buyers who have the correct valuations of each neighborhood in mind, and non-shocked buyers' valuations do not fully incorporate the costs associated with flood risk. This could be due to low salience, which has been shown to have a non-trivial welfare effect in the context of sales taxes (Taubinsky and Rees-Jones 2018). In this interpretation, it is the non-shocked buyers who are making

Table 1.7 · Sorting model estimates

	(1)	(2)	(3)	(4)	(5)	
	Utility terms			**Implied dollar value**		
	500-yr zone	no flood zone	α_p	500-yr zone	no flood zone	**IV method**
Full sample						
Destin	-0.132	-0.162	-0.071	-909	-1632	
	(0.103)	(0.092)	(0.814)	(1080)	(1675)	BT
Jacksonville	0.167	-0.371	-0.078	2354	-5036	direct
	(0.249)	(0.137)	(0.02)	(3609)	(2238)	
Miami	-0.071	-0.099	-0.137	62	63	BT
	(0.038)	(0.042)	(1.821)	(207)	(262)	
Naples	0.124	0.214	-0.024	7970	13800	direct
	(0.058)	(0.034)	(0.013)	(11813)	(21777)	
Tampa	-0.002	0.153	-0.067	33	2330	direct
	(0.14)	(0.083)	(0.009)	(2200)	(1345)	
Matched sample						
Destin	-0.311	-0.257	-0.018	-15661	-12819	direct
	(0.138)	(0.123)	(0.028)	(36009)	(35126)	
Jacksonville	0.083	0.011	-0.025	7361	1466	direct
	(0.096)	(0.069)	(0.012)	(18999)	(11367)	
Miami	0.055	-0.076	-0.268	378	-586	BT
	(0.052)	(0.049)	(0.558)	(349)	(375)	
Naples	0.083	0.070	-0.024	3933	3675	direct
	(0.074)	(0.038)	(0.01)	(3739)	(3489)	
Tampa	-0.026	-0.061	-0.004	-515	-1236	direct
	(0.12)	(0.078)	(0.006)	(2344)	(1650)	

Results from the sorting model estimated on data from each MSA. The α_p coefficient captures the marginal disutility associated with an extra $1,000 in annual rent. I combine the results in utility terms from Columns 1 and 2 with the estimates of α_p to calculate the implied dollar valuation of each amenity (Columns 4 and 5). Two IV strategies, the “direct” strategy and the Bayer-Timmins strategy (“BT”), were implemented for each sample, and the one with the more precise estimate was employed (see Section 1.5.2 for more details). Standard errors are boostrapped with 50 replications.

choices in a suspect choice environment, and their revealed preferences should not form the basis of the welfare calculation. Under this assumption, I can estimate the value of the disaster as an information disclosure about the risks of flood-prone areas:

$$\mathrm{cv} = \frac{1}{\gamma}\left[\log \sum_j e^{v_j^c} - \left(\log \sum_j e^{v_j^m} - \sum_j \pi_j^m (v_j^c - v_j^m)\right)\right]. \tag{11}$$

The absolute value of the compensation variation given by this equation will necessarily be the same as Equation 10, but the sign will be opposite. My analysis cannot determine which of these perspectives is right, although future research that seeks to measure this re-sorting process directly – are shocked buyers more or less likely than others to buy a new home within Florida in the years after their first purchase? – could help inform that question.

I estimate the compensating variation of flood exposure using the Naples model run on the matched sample from Table 1.7 above and find a value of \$425. This estimate suggests the average homebuyer is \$425 better off (or worse off, depending on assumptions about which group is mistaken) thanks to the alternative choices they made in light of their flood experience.

1.6 Amenity erosion: methodology and results

Based on my results in the previous sections, it seems that retreat from the water as an adaptive response to exposure to flood events is a real phenomenon. But the evidence above leaves many questions unanswered. Does this retreat reflect a major reinvestment of real estate capital with reverberations in coastal markets? Do non-shocked buyers enter the market to buy up coastal properties when demand is soft, or is the retreat phenomenon powerful enough to depress home prices in the retreat zone?

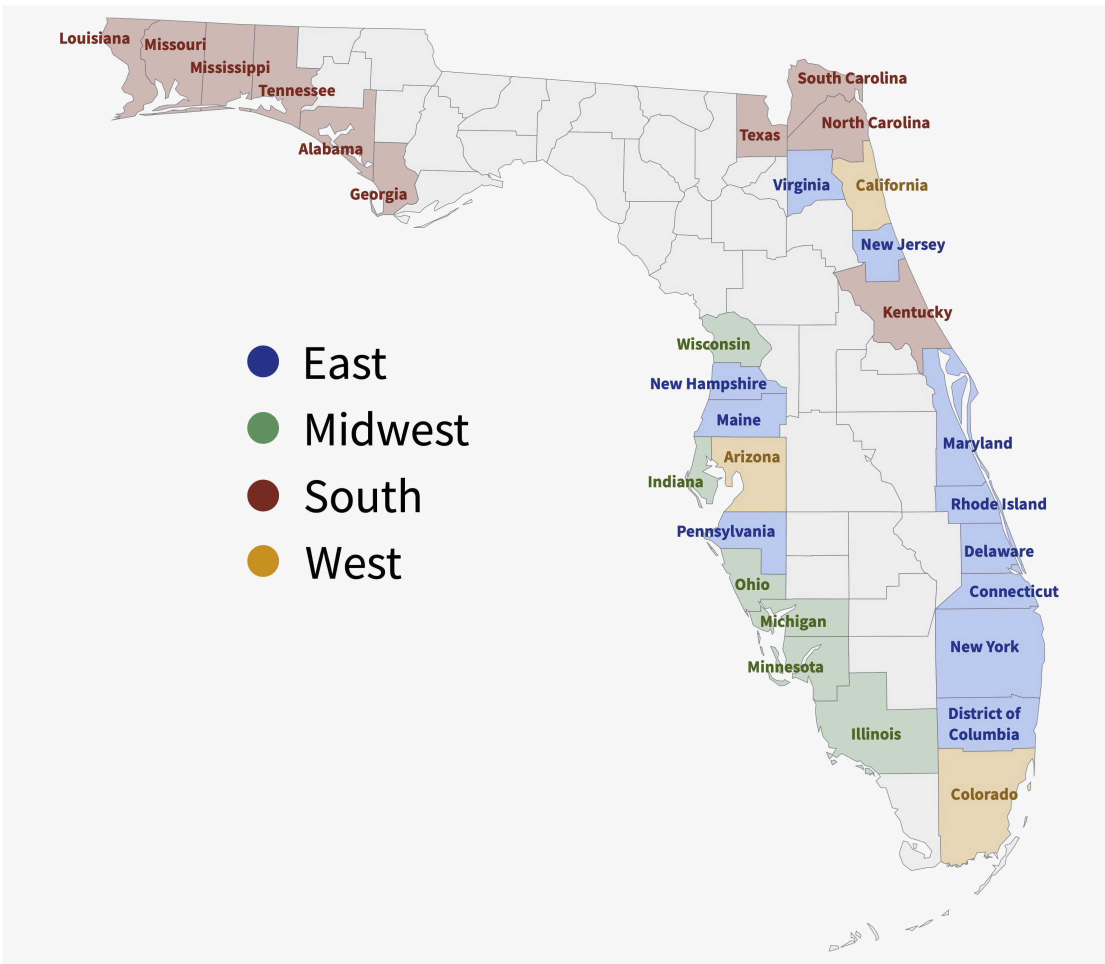
Proximity to water is ordinarily a highly valuable recreational and aesthetic amenity (Brown and Pollakowski 1977; Milon, Gressel and Mulkey 1984; Lansford and Jones 1995). However, this amenity naturally bundles water access together with increased flood risk. The relative weights attributed to flood risk and proximity services by buyers may, of course, change over time. To the extent that the marginal market participant become more worried about flooding, proximity to water should become less appealing, and may even become a disamenity.

This reasoning leads to my second hypothesis: the presence of disaster-shocked buyers will decrease, or erode, the equilibrium amenity value of proximity to the coast. This could occur because of shocked buyers' altered preferences, selective attrition from coastal properties, or both. The proximity to water amenity, for example, might become less valued in the market if fewer people are there to bid up homes near the water, or if disaster-shocked buyers decide to trade water proximity for other amenities, like a bigger home or an in-ground pool. This hypothesis is testable because markets vary in their share of disaster-shocked buyers, as measured by the disaster pressure index (see Figure 1.5).

Figure 1.11 highlights the source of identification in this analysis. While buyers from every part of the country enter Florida housing markets, there are some differences in the regional composition of buyers in different parts of the state. The figure shows that MSAs in the Florida panhandle are dominated by buyers from nearby southern states, Gulf Coast counties in southwestern Florida tend to have more Midwestern buyers, and the Atlantic coast counties are favored by East Coast buyers. Note that this map does not list the most common origin state in each county, which would be New York or Georgia in almost every case, but the origin state most distinctively linked to that county. This slight regional sorting among buyers is stable over time (see Appendix Figure A4), and allows for idiosyncratic changes in disaster pressure over time in different MSAs. If a series of floods hits the Midwest particularly hard one summer but leaves New England alone, there will be differing levels of disaster pressure impact in different parts of Florida.

Figure 1.11 · Most distinctive origin state for each Florida destination county

A map displaying the most distinctive buyer origin state for each Florida county, color coded by origin state region. While New York or Georgia is the single most popular origin state in nearly every counties, each county and region has its own distinctive catchment area across the country. Regional differences in origin between the panhandle, the Gulf Coast, and the Atlantic coast are very apparent. See Appendix A4 for a description of how each county's most distinctive state is determined.



1.6.1 Model specification

To test this hypothesis – that random variation in disaster pressure can drive changes to the real estate market in equilibrium – I adopt a standard hedonic price model framework. Following Freeman, Herriges and Kling (2014), I posit a utility function for a prospective buyer i in coastal Florida real estate markets considering purchasing home j:

$$U_i = u_i(X, Q_j, S_j, N_j), \tag{12}$$

where Q_j is a vector of the environmental characteristics of home j (such as proximity to water), S_j is a vector of structural characteristics (such as square footage), N_j represents neighborhood characteristics, and X is a composite good with price $= 1$. I assume that each real estate market participant purchases at most one home and that the market is in equilibrium, meaning that all participants select the home that maximizes their personal utility function at prevailing prices. This implies that the sale price of a property is a function of property characteristics with the following form:

$$R_{h_j} = R_h(Q_j, S_j, N_j). \tag{13}$$

Equation 13 is the hedonic price function. Assuming the individual maximizes utility U_i subject to a fixed budget constraint, the first-order condition for optimality with respect to element q of Q_j is:

$$\frac{\partial U_i/\partial q}{\partial U_i/\partial X} = \frac{\partial R_h}{\partial q}. \tag{14}$$

Equation 14 states that an individual's marginal willingness to pay for characteristic q (left-hand side) is equal to the implicit marginal price of q in the hedonic function (right-hand side), which can be estimated econometrically. I estimate a version of the hedonic price function from Equation 13 with the following specification:

$$\log(P_{ijt}) = \alpha + \beta_x \mathbb{X}_i + \beta_w \mathbb{W}_i + \gamma Z_i + \delta t_i + \epsilon_i, \tag{15}$$

where P_{ijt} represents the price of property i sold in Florida MSA j in month t, $\mathbb{X}_i$ is a vector of structural characteristics like home size, $\mathbb{W}_i$ is a vector of characteristics pertaining to proximity to the water and, Z_i and t_i are ZIP code and month of sample indicators respectively, and ϵ_i is an error term. I then augment Equation 15 to include an interaction between each characteristic and disaster pressure:

$$\log(P_{ijt}) = \alpha + \beta_{x,1}\mathbb{X}_i + \beta_{x,2}[M_{ij} \times \mathbb{X}_i] + \beta_{w,1}\mathbb{W}_i + \beta_{w,2}[M_{ij} \times \mathbb{W}_i] + \gamma Z_i + \delta t_i + \epsilon_i, \tag{16}$$

where M_{ij} is the disaster pressure for market j in month i, in percentage points. The $\beta_{x,1}$ and $\beta_{w,1}$ coefficients embody the amenity premium in the absence of shocked buyers (i.e. when $M_{ij} = 0\%$), while the $\beta_{x,2}$ and $\beta_{w,2}$ reflects the influence of disaster-shocked buyers on the amenity value. My hypothesis implies that the $\beta_{w,2}$ coefficient for each characteristic will be of the opposite sign of the corresponding $\beta_{w,1}$ coefficient, indicating that the greater the presence of disaster-shocked buyers in a particular market, the lesser the marginal willingness to pay for water proximity at that time and place.

1.6.2 Results from hedonic models

My hedonic model estimates are displayed in Table 1.8. I display estimated main effect and interaction effect coefficients for each of four main structural characteristics (square footage, year built or last renovated, pool indicator, and garage indicator) and seven water proximity characteristics (frontage, four flood zone indicators, and two sea-level-rise zone indicators). Columns 1 through 3 display estimates model estimated on the full sample of homes from Table 1.1, Column 2, aggregating across time and across MSAs. Recall that this is considerably larger than the sample of homes used to estimate the count models in Table 1.4, as I am no longer restricted to properties with known origin – I can include every home in the sample to get the most precise and valid estimates of marginal value possible. I do restrict the sample to MSAs with no local shock in the past 24 months, so the

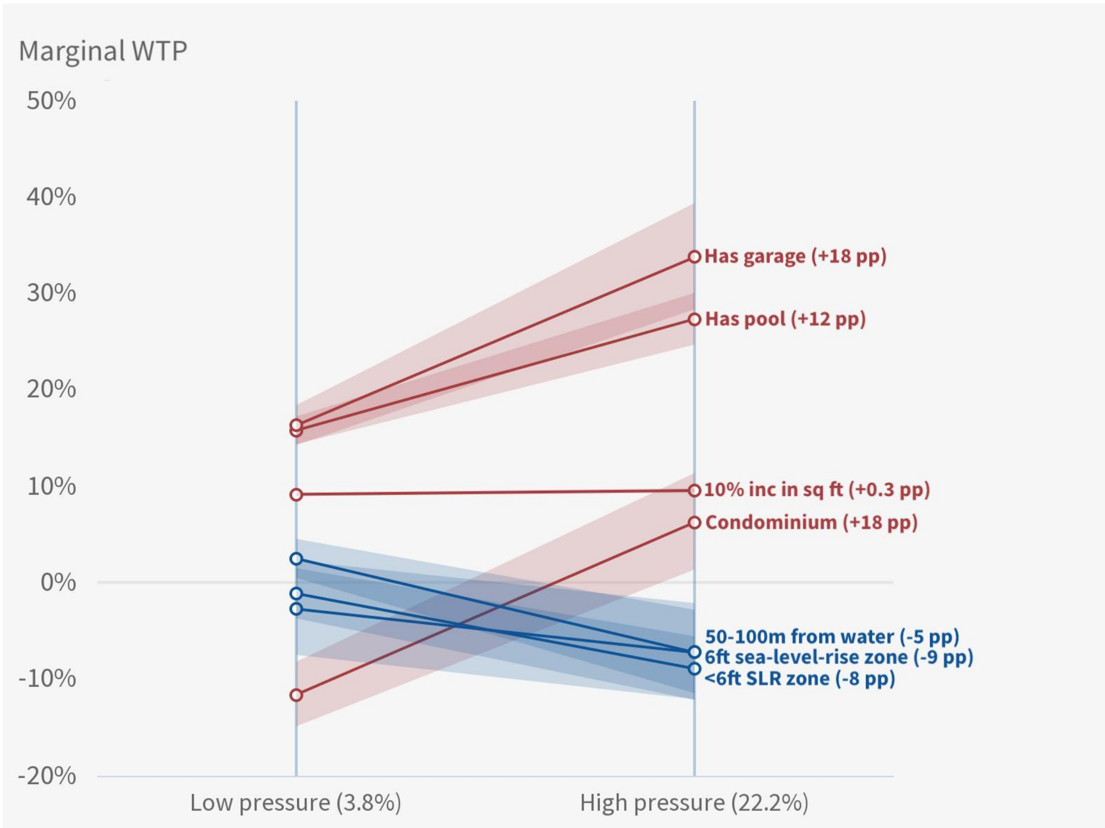
Table 1.8 · Statewide hedonic price models

	(1)	(2)	(3)	(4)	(5)
Hydrology controls:	Distance bins	Flood zones	SLR zones	Distance bins + flood zones	distance bins + SLR zones
Property characteristics ($\beta_{x,1}$)					
Square footage (log)	0.871***	0.914***	0.915***	0.874***	0.874***
	(0.0120)	(0.0122)	(0.0123)	(0.0120)	(0.0120)
Year built/renovated	0.00394***	0.00375***	0.00375***	0.00398***	0.00396***
	(0.000392)	(0.000371)	(0.000369)	(0.000393)	(0.000391)
Pool indicator	0.128***	0.132***	0.131***	0.129***	0.129***
	(0.00924)	(0.00894)	(0.00901)	(0.00921)	(0.00929)
Garage indicator	0.121***	0.146***	0.147***	0.122***	0.123***
	(0.0115)	(0.0110)	(0.0110)	(0.0116)	(0.0116)
Hydrology main effects ($\beta_{w,1}$)					
Frontage	0.142***			0.197***	0.185***
	(0.0328)			(0.0317)	(0.0320)
VE (100-yr flood + storm surge)		0.245***		0.253***	
		(0.0665)		(0.0632)	
A (100-yr flood plain)		0.0313**		0.0353**	
		(0.0151)		(0.0155)	
AH (100-yr shallow flooding)		0.0428***		0.0468***	
		(0.0150)		(0.0146)	
X (500-yr floodplain)		0.0392***		0.0445***	
		(0.0123)		(0.0123)	
< 6-ft SLR zone			0.00234		0.00550
			(0.0184)		(0.0186)
6-ft SLR zone			0.0398***		0.0456***
			(0.0135)		(0.0135)
Hydrology interaction effects ($\beta_{w,2}$)					
Frontage	-0.00616***			-0.00190	0.000656
	(0.00157)			(0.00161)	(0.00161)
VE (100-yr flood + storm surge)		0.00896***		0.00496	
		(0.00341)		(0.00321)	
A (100-yr flood plain)		0.00107		0.000366	
		(0.000905)		(0.000965)	
AH (100-yr shallow flooding)		-0.00525***		-0.00556***	
		(0.00115)		(0.00116)	
X (100-yr flood plain)		-0.00207**		-0.00249***	
		(0.000818)		(0.000844)	
< 6-ft SLR zone			-0.00297**		-0.00445***
			(0.00118)		(0.00120)
6-ft SLR zone			-0.00417***		-0.00542***
			(0.00141)		(0.00142)
3,068 Census tract FEs	✓	✓	✓	✓	✓
Quarter-of-sample FE	✓	✓	✓	✓	✓
Observations	2,342,896	2,342,896	2,342,896	2,342,896	2,342,896
r^2	0.407	0.405	0.404	0.405	0.406

OLS estimates with robust standard errors clustered at the census tract in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is the log of sales price in nominal dollars. Frontage properties are those within 50 m of the water. Disaster pressure is the percentage of buyers in an MSA-month whose home counties experienced a flood event in the 24 months leading up to the transaction. The five specifications shown here differ only in the set of hydrology controls included.

Figure 1.12 · Change in amenity value induced by rising disaster pressure

Taking results from the model displayed in Table 1.8, Column 5, I estimate the amenity value associated with each characteristic under two scenarios: low disaster pressure and high disaster pressure (these values are the 5th and 95th percentiles of disaster pressure in the data). I display those point estimates along with 95% confidence bands. Non-water amenities like garages and pools are valued more highly in equilibrium when disaster pressure is high, while the relative appeal of water proximity falls.



maximum value of the disaster pressure index in the data is approximately 30%.

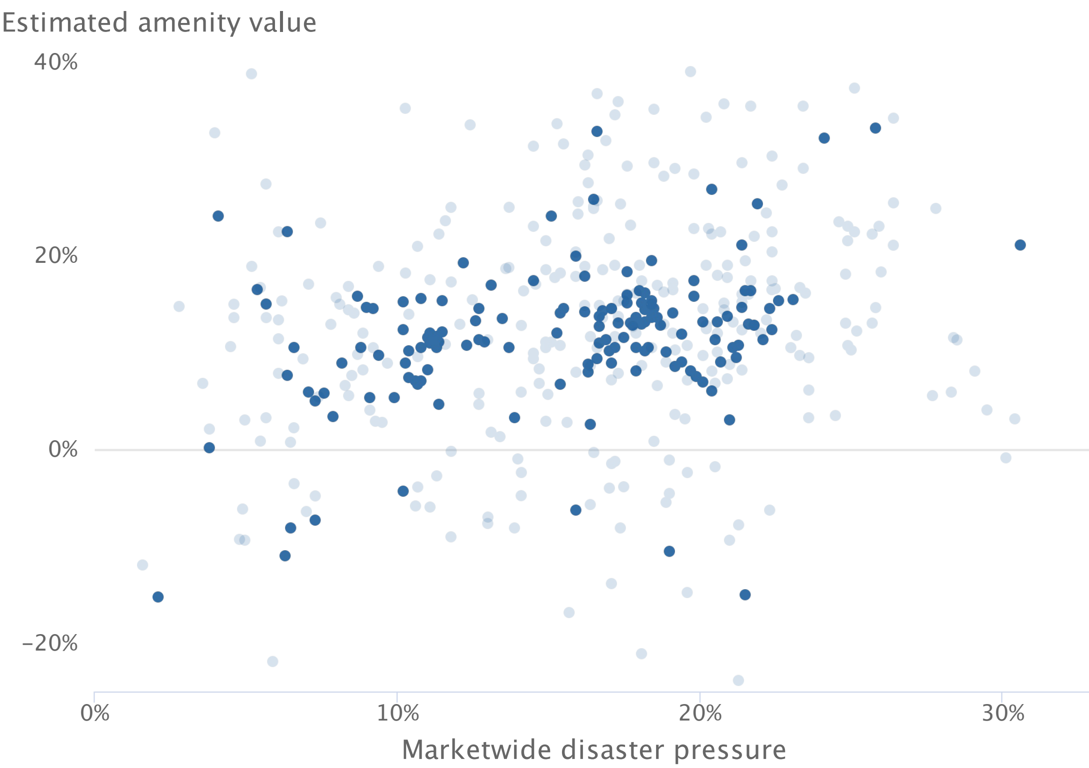
The estimates for $\beta_{w,1}$ indicate that coastal proximity, as expected, is a valuable amenity in these markets. Frontage properties command significantly higher prices than similar properties just inland. Meanwhile, properties located in flood zones sell for a 3-5% premium relative to homes at the same distance to the water, but outside flood zones. While flood zone status does not immediately seem like an appealing characteristic, it likely proxies for favorable amenities like ease of access to the waterfront that are not perfectly captured by coastal distance and elevation data. Likewise, sea-level-rise zone designations may capture information about water access in its own way, and homes in the 6-ft SLR zone also sell at approximately a 5% premium. In both cases, these coefficients are intact in models where I control separately for distance, as in Columns 4 and 5.

My hypothesis is that these coefficients, which indicate buyers' marginal willingness to pay for each characteristic under the standard hedonic assumptions, will decline or reverse when more disaster-shocked buyers enter the market. Specifically, I predicted that the coefficients $\beta_{w,2}$ to be of the opposite sign as the corresponding $\beta_{w,1}$ coefficients, and that is indeed what the models show.

Figure 1.12 displays the results from the specification in Column 4 of Table 1.8 in graphical form, including information on $\beta_{x,2}$ coefficients that were not presented in the table. I estimate the combined coefficient on $\beta_{.,1}$ and $\beta_{.,2}$ under two disaster pressure scenarios, and display those point estimates along with 95% confidence

Figure 1.13 · Hedonic meta-analysis of flood zone disamenity

Results from 374 separate hedonic regressions conducted at the MSA-quarter level, using a specification analogous to the model displayed in Table 1.8, Column 2. Point estimates of the marginal amenity value of non-flood-zone location are plotted against the average MSA-quarter disaster pressure (values on both axes adjusted for MSA averages over time). The tercile of estimates with the tightest standard errors are highlighted. Taken together, the results suggest a positive relationship between disaster pressure (which I argue is quasi-randomly assigned after adjusting for MSA) and the equilibrium market value of flood zone location within census tract.



bands. Positive slopes indicate that amenity value increases when more disaster-shocked buyers enter, while negative slopes indicates amenity erosion. In keeping with my hypothesis, amenities associated with flood risk (lines) tend to erode when more shocked buyers enter, while other amenities (red bars) tend to rise in value.

This provides some additional evidence that the retreat effect identified above reflects underlying changes in buyers' marginal willingness to pay for various amenities. When more disaster-shocked buyers participate, they evidently bid up the prices of inland houses with pools and garages relative to the price of smaller homes closer to the beach. This runs counter to the proposition that disaster-shocked buyers are simply decreasing spending across the board, or that the retreat effect is not economically meaningful at the level of the whole market.

The underlying assumptions that allow me to infer marginal willingness to pay from the coefficients of the hedonic price function require that all data pertains to a single market where participants could, in principle, select any home. This assumption may not hold for a sample including 15 different MSAs over 13 years, as some buyers surely face geographic and chronological constraints on their purchases, so results from the full sample are potentially not interpretable in the customary way (Freeman, Herriges and Kling 2014). I include MSA-level and quarter-of-sample fixed effects in the model, but this only admits a level shift in prices across MSAs, not differential response to an influx of shocked buyers.

To address this concern, I repeat this analysis while segmenting the sample by MSA and calendar quarter and estimating the model on each market separately, using a specification analogous to the model displayed in

Table 1.8, Column 2, including flood zone bins and census tract fixed effects, but no separate distance control. Figure 1.13 plots the results of 374 separate hedonic regressions. In each case, I take the point estimate for the amenity value associated with being located outside a flood zone, and plot it against the local disaster pressure in MSA j at quarter q, averaged across the three months in each quarter. As some estimates are relatively noisy, I highlight the estimates with the lowest tercile of standard error magnitude. While I do not attempt a precise statistical test on this collection of estimates, the overall positive relationship is consistent with what I find in the pooled hedonic regressions.

1.6.3 Property value erosion calculation

I proceed to use the estimates from the pooled hedonic model to estimate the aggregate effect of disaster-shocked buyers on the total valuation of the coastal amenity in each market. To do so, I first use the fitted hedonic models to estimate the value of water proximity capitalized into each home, by comparing the predicted value of each home according to my estimate of the hedonic model R_h (Equation 16) with its actual characteristics, to the predicted value of a hypothetical home with the same structural characteristics $\mathbb{X}_i$ but modified water proximity characteristics $\mathbb{W}'_i$:

$$A_i = \widehat{R_h}(\mathbb{X}_i, \mathbb{W}_i, M_{jt}, z, j, t) - \widehat{R_h}(\mathbb{X}_i, \mathbb{W}'_i, M_{jt}, z, j, t), \tag{17}$$

where A_i is the estimated coastal amenity value for property i and $\widehat{R_h}$ is the predicted value based on my estimation of Equation 16. Roughly speaking, I "move" each property to a hypothetical position 5 kilometers from the coast and outside of flood zones and sea-level-rise zones – while retaining its structural characteristics – and then note the decline in predicted sales price. Naturally, any effort to interpret this difference as the value of the coastal amenity must be undertaken with caution (Brown and Pollakowski 1977; Loomis and Feldman 2003); I discuss limits to interpretability below.

I next simulate A_i under two scenarios with 4% and 22% of buyers shocked:

$$A_{i,4} = \widehat{R_h}(\mathbb{X}_i, \mathbb{W}_i, M = 4, z, j, t) - \widehat{R_h}(\mathbb{X}_i, \mathbb{W}'_i, M = 4, z, j, t), \tag{18}$$

$$A_{i,22} = \widehat{R_h}(\mathbb{X}_i, \mathbb{W}_i, M = 22, z, j, t) - \widehat{R_h}(\mathbb{X}_i, \mathbb{W}'_i, M = 22, z, j, t). \tag{19}$$

These values correspond to the 5th and 95th percentiles of the conditional distribution of the disaster pressure index for observations with no local flood event in the past 24 months, respectively. The difference in these quantities provides a measure of how much the value of the coastal amenity capitalized into each property declines when significantly more disaster-shocked buyers enter the market.

I use the estimates displayed in Table 1.8, Column 4 to answer the following two questions: how big in dollar terms is the coastal amenity capitalized into the value of each home in these markets, and how does this amount *change* when disaster-shocked buyers arrive en masse? Following the procedure outlined above, I use the estimates to infer how much each property value would be reduced if that property were moved to a location 5+ km from the water, outside any flood zones and sea-level-rise zones, but otherwise were left unchanged.

As I argue in the previous section, the hedonic model – and hence my estimate of coastal amenity value – depends critically on the mindset of the buyers in the market and specifically how many of them have recently experienced disaster shocks. I posit two possible values for disaster pressure, 4% and 22%, and compare the average estimated dollar value of coastal amenity in each market under each scenario. I construe property value reduction as the difference in the amount of coastal amenity capitalized into the price of the home under each scenario.

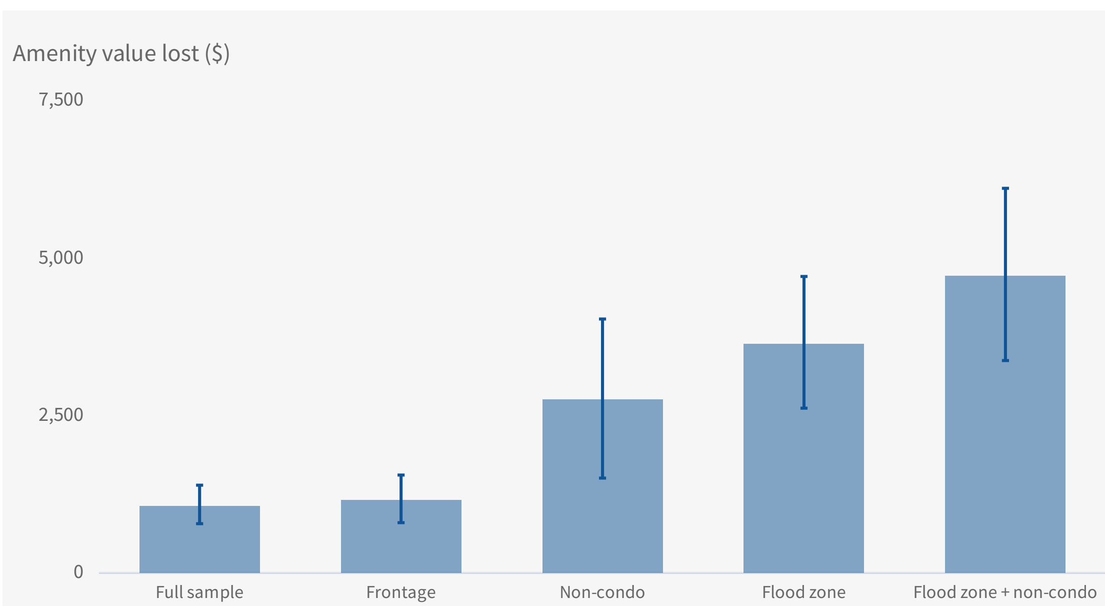
My estimates show that the average home in the sample loses approximately $1,000 of value (Table 1.9). This would suggest that flood events in far-off northeastern and Midwestern counties one summer can significantly impact property values in Sarasota and Daytona Beach the next spring, and that the adaptive retreat pattern found in the counts analysis has real economic bite. The harm is concentrated in certain submarkets where the coastal amenity is a larger part of the typical home's value; I estimate that the average frontage property loses

Table 1.9 · Amenity erosion with influx of shocked buyers

Sample	Mean sales price ($k)	Lost amenity value ($)	Percentage value loss
Full sample	222	1,082 (158)	-0.5%
Frontage (<50m) properties	463	2,764 (647)	-0.6%
Non-condo properties	223	1,173 (192)	-0.5%
Flood zone properties	278	3,655 (532)	-1.3%
Flood zone + non-condo	249	4,734 (699)	-1.9%

Results from the amenity erosion calculation described in Section 1.6.3. Bootstrapped standard errors for the lost amenity value estimates are in parentheses.

Figure 1.14 · Coastal amenity erosion across different submarkets
The estimated decrease in the market valuation of coastal amenity associated with a hypothetical large increase in disaster pressure. This figure displays the same results as Table 1.9. See Section 1.6.3 for further explanation of how these figures are calculated.



nearly $2,800 in value and the worst-hit group of homes (non-frontage, non-condo properties in a flood zone) lose $4,700, or nearly 2% of their total value on average.

Limiting the reliability of this exercise is the relatively strict assumption that the inverse demand curve can be recovered from the point estimates of the hedonic price function (Freeman, Herriges and Kling 2014). As highlighted by Palmquist (1992) and Loomis and Feldman (2003), extrapolating the marginal willingness to pay estimated in hedonic models across non-marginal changes (for example, comparing a house 100 meters from the water to a house 10 km from the water) is not generally reliable. Palmquist (1992) argues that such calculations are interpretable as welfare changes only when the number of properties affected by an amenity is low relative to the size of the market, which is not the case in this context. These estimates are probably best viewed as rough indications of the order of magnitude of the true quantity (Brown and Pollakowski 1977). Even so, value losses on this order of magnitude multiplied by hundreds of thousands of properties in a given market would certainly represent a major economic shock in the aggregate.

1.7 Conclusion

In this chapter, I take advantage of a rich dataset of Florida real estate transactions, including information on buyers' counties of origin, to estimate the effect of disaster shocks on shocked residents' home purchase patterns, and more broadly on the real estate markets they enter. I find that shocked home buyers respond by differentially exiting the coastal Florida real estate market, and by retreating away from the coast when they do enter.

This novel finding of climate adaptation in real estate markets, and evidence that it significantly alters the market equilibria when shocked buyers are present in significant numbers, is an important instance of climate adaptation with concerning implications for people who hold trillions of dollars worth of Florida coastal real estate. However, the transience of the effects I find suggests that even major, high-profile disasters will not permanently alter the trajectory of climate adaptation.

In future research, I hope to learn whether more intensity or repetition of disasters can lead to a more permanent state of shock. I also seek to answer the question of whether the influx of shocked people has the potential to change minds in the home community: do shocked buyers slowly but surely change the minds of native Floridians, or is there hesitation to move to the water forgotten after one or two years? Finally, combining this approach with survey data on people relocating to Florida could shed light on the mechanisms at work, and indicate to what extent altered beliefs are responsible for the major effects I observe.

Chapter 2

Risk aversion in the field: evidence from an employee rewards program

2.1 Introduction

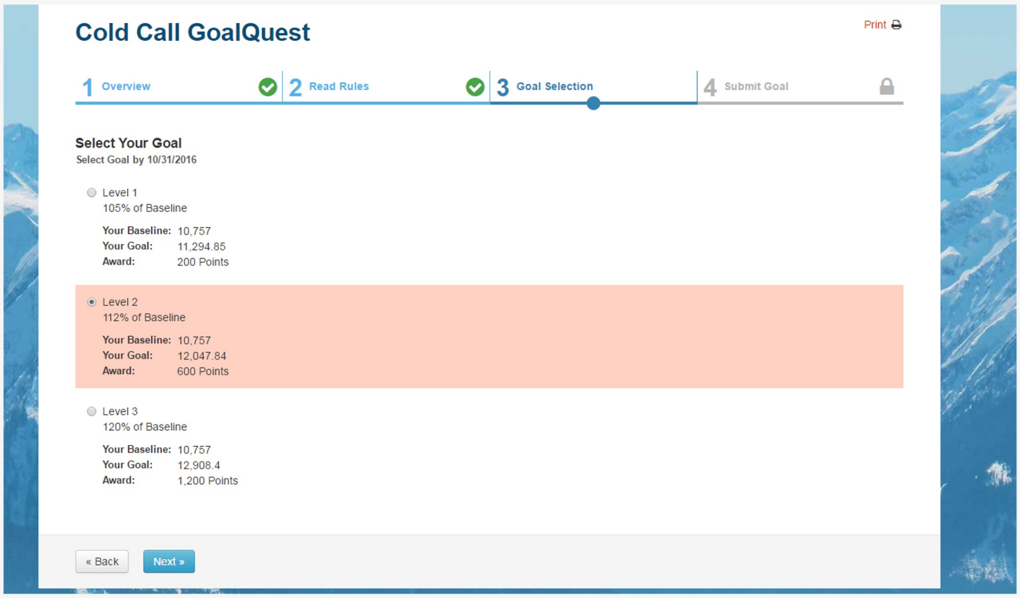
Economists have long sought to understand the motives for financial risk-taking. Clarifying preferences for risk has profound implications for economic theory, estimates of welfare, and the optimal design of programs and policy. As examples, such insight should, in theory, inform how policymakers regulate financial markets or design entitle programs, how firms price insurance plans and design contingent contracts for employees, how individuals optimize their asset allocations, and how economists evaluate the welfare effects of policies and programs involving financial risk. At first glance, individuals appear to exhibit substantial risk aversion: that is, they avoid higher, but riskier, financial outcomes in favor of lower, less risky, outcomes. Within economics, standard theory attributes risk-aversion among fully-informed, utility-maximizing individuals, to the diminishing marginal utility of wealth, implied by the concavity of their utility function. However, recent research has advanced several departures from the standard model that could explain risk aversion through alternative channels. These channels include systematically biased beliefs, non-linear decision weights, aversion to losses, disappointment or regret in the context of gain-loss utility (Barseghyan et al. 2018). Risk aversion could also emerge from non-standard decision processes involving heuristic choice, emotion, or hormones (e.g., Kusev et al. 2017).

An expansive empirical literature has sought to examine measure individual appetite for risk and to understand its motives. In the lab, researchers have amassed considerable evidence as to the behavior of individuals in response to lotteries involving small-to-medium sized stakes. In the field, researchers have investigated risky choice in more naturalistic settings primarily involving insurance, betting, and game shows. Clarifying motives for risk-taking in the field, however, is practically complicated by potential confounds such as unobserved bias in beliefs (e.g., betting markets, insurance), imperfect understanding of choice (e.g., insurance), or limited generalizability (e.g., game shows).

We overcome several of these challenges by analyzing an unusually rich dataset describing the decisions and beliefs of employees in the context of a simple, all-or-nothing, employee goal-rewards program, called GoalQuest (GQ). The program was conceived by BI Worldwide (BIW), a US-based consulting firm specializing in the design and administration of employee/consumer engagement programs that leverage principles of behavioral science. More specifically, employee participants in GQ were asked to self-select a productivity goal for the one- to three-month duration of the program from a menu of three options, personalized based on performance during a pre-program control period. To encourage employees to choose aggressively, goal thresholds typically increased linearly (e.g., Goal 1: 100 units, Goal 2: 110 units, Goal 3: 120 units) while all-or-nothing rewards increased non-linearly (e.g., Goal 1: \$100, Goal 2: \$300, Goal 3: \$600). As a result, for most employees, Goal 3 provided the highest reward in expectation.

Figure 2.1 · GoalQuest goal selection webflow

A screenshot from the web portal that GoalQuest participants use to select goals. Goal levels are customized and explicitly compared to baseline productivity. Employees separately receive information about the effective exchange rate between award points and dollars at the GQ marketplace, and many have prior experience valuing the points in the context of other programs conducted by BIW.



Our primary evidence reflects the goal choices, and beliefs regarding goal attainment, of 20,133 employees who participated in 34 distinct GQ programs administered across 18 large North American (primarily US) firms from 2016 to 2019. These employees stood to earn \$9.4 million in potential rewards through the program. We corroborate our results with additional data describing the goal choices and productivity, but not beliefs, of another 19,221 employees. To organize potential explanations for conservative goal choice, we outline a simple theoretical framework in which, at baseline, a fully-informed, utility-maximizing, risk-neutral, employee must select either a high or low goal associated with an all-or-nothing reward. The framework then introduces departures from this baseline – biased beliefs, non-linear decision weights, and gain-loss utility – informed by the literature. We proceed to characterize the optimality, and conservativeness (and aggressiveness), of employee goal choice relative to benchmark models emerging from the framework. Finally, to further explore additional mechanisms, and to examine potential confounds, we supplement our analysis from the field with online experiments simulating an abbreviated GQ program in the context of incentive-compatible effort tasks.

Several aspects of the research setting make it particularly attractive for understanding how people engage risky financial decisions in the field. A first distinguishing feature is the high generalizability of the setting. One factor that contributes to the generalizability is the diversity of the sample. Across GQ programs, we observe a diversity of age, occupation, industry (e.g., communications, health care, manufacturing, financials, consumer discretionary, consumer staples), geography, industry experience, and salary. The representativeness of the sample is alternatively conveyed by the popularity of GQ programs across the US. While our data is limited to recent programs that adopted an enhanced enrollment process, our programs our observably similar to the historical universe of programs which have been collectively been administered to over 1 million employees across a significant fraction of Fortune 500 firms. And in contrast to many economic contexts that have attracted research attention, such as betting markets or game shows, there is minimal selection into, or out of, the sample (BIW markets the program as having a participation rate in excess of 98 percent). The generalizability of the setting is also reflected in the wide-range of financial stakes faced by employees. We note that the estimated value of the rewards in our data range from tens of dollars to $2,813, a range that overlaps with many other economic decisions of interest to economists. A second distinguishing feature of the setting is that, despite the diversity reflected in the composition of decision-makers and in economic stakes, employees render their decisions from a highly standardized and simple choice menu. Unlike complicated economic settings such as the choice of insurance plans, employee participants in GQ were intended, by design, to have a clear understanding of program rules and financial stakes (a presumption we confirm in the lab). Finally, our partnership with BIW led to the creation of an enhanced enrollment process in which employees directly reported their beliefs as a part of program onboarding. As a result, our analysis leverages administrative data not just on goal choice and productivity, but on contemporaneously elicited beliefs of goal attainment. Ultimately, we see this setting as offering a litmus test into the motives of choice under financial risk and uncertainty across diverse contexts of economic interest.

We describe several findings from our analysis of employee data. Our first finding is to document substantial risk aversion in the goal choices of employees. Assuming employees have rational ex-ante beliefs of goal attainment, which we estimate using a procedure borrowed from the economic literature on insurance, we conclude that 49 percent of employees selected an overly conservative goal relative to the expected value (EV) maximizing benchmark (most conservative goal choice involved the selection of Goal 1 or 2 rather than the EV-optimal Goal 3). For employees succeeding in attaining the low-goal threshold, conservative goal choice resulted in an average expected foregone reward of $139. Relative to their actual realized productivity (perfect information), the average foregone reward associated with conservative goal choice was $175. The excess conservatism of employees, as judged from the rational expectation or perfect information benchmark, persisted in decisions with large financial stakes and among highly tenured employees (10+ years of experience).

Our second finding is to show that conservative goal choice cannot be explained by the diminishing marginal utility of employee wealth, the central explanation for risk aversion in the standard model. Specifically, we show the descriptive accuracy with which the expected utility benchmark can explain goal choice does not meaningfully increase under any plausible degree of risk aversion as modeled using the CARA utility function. While increasing assumed risk aversion moderately reduces the share of seemingly conservative choice, it increases the share of seemingly aggressive choice by roughly the same magnitude (i.e., for most employees, as assumed risk aversion increases, the optimal choice shifts from Goal 3 to Goal 1). Even a hyper-flexible benchmark model of heterogeneous risk preferences that tags a goal choice as explained if it can be rationalized by any plausible set of risk preferences still characterizes 45 percent of employee choices as sub-optimal.

Our third finding is to show that conservative goal choice cannot be explained through biased beliefs. In our theoretical framework, we describe how one explanation for conservative goal choice is if otherwise utility maximizing employees were under-confident about attaining a higher, relative to a lower, goal (or, equivalently, were relatively over-confident about attaining a lower goal). Leveraging access to specific employee beliefs of attaining each goal, we show that, on average, employees were substantially overconfident about the likelihood of achieving all of the goals, and were relatively overconfident about reaching the higher goal. Accordingly, we show that shifting from a benchmark model assuming rational expectations to one using subjective belief of goal attainment does not meaningfully change the characterization of choice across the range of plausible risk preferences. Ultimately, a model of subjective utility maximization, assuming reasonable degrees of risk aversion, implies that nearly one-half of employees, including those facing decisions with rewards falling in the highest quartile and those in the highest tenure category, choose sub-optimally.

Our next finding is to show that the incorporation of two prominent, behaviorally-informed, departures from the expected utility framework – gain-loss utility and non-linear decision weights – does not substantially improve the predictive accuracy of the benchmark model. Regarding the former, given the precedence in the literature for explaining conservative financial choice through loss aversion in the context of reference-dependent preferences, we characterized employee goal choice using benchmark models featuring gain-loss utility. We considered a range of prospect-independent (e.g., Goal 1, Goal 2, Goal 3) and prospect-dependent (e.g., Reward of Chosen Goal, Expected Value of Chosen Goal, Counterfactual Reward) reference points. The set of candidate reference points reflected the theorized role of expectations (Kőszegi and Rabin 2006), disappointment (Gul 1991; Loomes and Sugden 1986), and counterfactual regret (Loomes and Sugden 1987). Across these potential reference points, as well as a range of plausible loss aversion parameters, and various convex combinations of gain-loss and consumption utility, we find that models of gain-loss utility explain, at most, a moderate additional share of goal choice: the most successful of these models increases descriptive accuracy from 50 to 59 percent. Regarding non-linear decision weights, while research has established the propensity of decision-makers to overweight low-likelihood outcomes and under-weight those with a high likelihood, we find that adopting a non-linear weighting function (e.g., Prelec 1998) does little to change our characterization of employee choice.

To generate additional evidence on mechanisms, and to rule out potential confounds, we designed and administered an experimental incentive-compatible rewards program, resembling GQ, in the context of an online effort task. The experimental paradigm permitted us to directly elicit/assess person-specific decision-making parameters (e.g., beliefs, risk preferences, loss aversion, taste for competition, emergency liquidity), confirm understanding of program rules, and observe multiple goal choices per subject across six strategically varying menus (for incentive compatibility, subjects were told that only one of their goal choices would be randomly selected to calculate their reward). The exercise, while stylized and involving far smaller stakes than in the field, produced the same general pattern of goal choice and substantial overconfidence about productivity as observed in the field.

We then proceeded to assess the predictive accuracy of a series of benchmark decision-models, beginning with a baseline model of risk-neutral, expected-utility maximization, with rational expectations. As in the analysis of field data, we find that one cannot adequately explain the goal choices of experimental subjects using the expected utility framework, even after incorporating plausible levels of risk aversion, subjective beliefs, non-linear decision weights, or gain-loss utility using the best-performing parameters and functional specification from the field. We also document the limited success of "fuzzy" utility models that allow for some computational error in the calculation, or comparison, of subjected expected values. Finally, we assess, and then also reject, two alternative mechanisms that have been suggested in the literature as potential explanations for menu-based choice involving contextual sorting (e.g., Kamenica 2008) and a taste for competition. The most descriptively successful model – composite gain-loss utility – explained only 18 percent of subject choices across all menus and 40 percent of subject choices across all-but-one menu.

Ultimately, the evidence across the field and lab implies that the standard expected utility framework, even after modification to include departures most seriously contemplated in the literature, has limited ability to explain the simple, yet consequential, choices in this context. These limits extend to the highest stakes settings we observe and to the most experienced employees in the sample. We conclude by describing alternative decision-making mechanisms that may help to explain risky financial choice in this setting. In particular, we propose a novel anxiety-based heuristic that offers a plausible explanation for the behavior, and beliefs, documented across the field and the lab. The heuristic is premised on the possibility that, for most employees, goal choice induces some anxiety. The heuristic stipulates that the decision to select the expected-value maximizing choice (typically Goal 3) might be overruled if the anxiety associated with choosing the high goal is sufficiently high. Our speculation as to the role of emotion in these decisions is consistent with the theorized role of anticipated emotion on economic decisions (Loewenstein and Lerner 2003) and more explicit consideration of how emotions affect risk-taking (e.g., Slovic 2004; Miu, Miclea and Houser 2008). The avoidant effect of anxiety on decisions has been established by considerable research in neuroscience and psychology, with evidence additionally suggesting that such avoidant effects are hard-wired (Hartley and Phelps 2012). Through an additional experiment using the same online goal-reward paradigm, we show that when the high goal induces sufficient anxiety, decision-makers retreat to a lower goal choice, despite their beliefs.

Collectively, we see our evidence as contributing to a growing literature seeking to better understand how individuals navigate decision-settings involving financial risk. Specifically, our analysis corroborates the pervasiveness of risk-aversion that other researchers have found in settings involving insurance choice, game shows, and betting markets (Sydnor 2010; Snowberg and Wolfers 2010; Hartley, Lanot and Walker 2014; Beetsma and

Schotman 2001). Our results build on this literature by showing that risk averse behavior extends to a highly diverse cross-section of individuals asked to make decisions across widely-varying financial stakes in a simple choice context.

We also contribute new insight into the motives that underlie financial risk aversion, given our access to data on employee choice and beliefs. A primary take-away is that risk aversion in the field may not reflect actual preferences for risk in the manner understood by the standard economic model, nor by common modifications to the model advanced in the literature. We suggest instead that risk averse decisions in this setting, and possibly other settings as well, emerge from non-standard decision processes involving an emotion-based heuristic. If true, a practical implication is that an understanding of the potentially non-standard decision-processes that govern how individuals engage financially risky choice may be critical in shaping the optimal design, and administration, of associated policies and programs. For example, these findings imply that policymakers and firms could potentially increase welfare by adopting choice environments, or decision tools, that discourage the use of emotion-based heuristics. Similarly, for economists, insight into underlying psychological motives may be a precursor for accurately assessing the welfare effects of policies in markets involving assets or insurance.

2.2 Background

2.2.1 Institutional background

GoalQuest®(GQ) is an incentive program conceived and administered by BI Worldwide (BIW), a private global consulting firm. The firm specializes in the design and delivery of a suite of proprietary programs that leverage principles from behavioral science (e.g., non-monetary rewards, goal-setting, personalization, symbolic recognition, lotteries, contests, communication, and feedback) to improve employee, channel partner, and consumer engagement. Founded in Minneapolis, Minnesota in 1950, the firm, as of 2021, had more than 25 sales offices across 9 countries and had self-reportedly engaged 6 million individuals across 144 countries through its various products. According to third-party estimates, as of 2021, the firm had approximately 1,500 employees and annual revenues between $500 million to $1 billion.

Described by BIW as the "world's only patented sales incentive design," GQ was designed to motivate employee productivity through self-selected performance goals tied to all-or-nothing non-monetary rewards. As of 2018, BIW had administered over 1,000 GQ programs to over 1 million participants at firms primarily in the United States, Canada, and Europe since its 2001 inception. While marketed as a sales incentive program, our data indicate that the program has serviced a significant share of employees engaged in customer service and retention (e.g., call centers) across an impressive diversity of sectors (e.g., communications, health care, manufacturing, financials, consumer discretionary, consumer staples). Typically, a GQ program engages hundreds to a few thousand employees with larger programs involving tens of thousands of employees. Employee participants range from newly hired to highly tenured and from staff workers to managers and supervisors.

2.2.2 Reward program overview

Across the wide-range of client firms, GQ boasts a uniform program structure, particularly since 2012, the year of our earliest data. From an employee's perspective, program participation entails three phases: enrollment/goal selection (and program marketing), a performance period typically lasting between 30 and 90 days, and, for those achieving their selected goal, reward redemption. During the initial phase, employees are asked to enroll in the program, and select their goal, by visiting an online portal and proceeding through a simple webflow.

The webflow itself consists of four phases: a program overview, an enumeration of program rules, goal selection, and goal confirmation (see Figure 2.1 for a sample screenshot). To select a goal, in most programs, employees are asked to select from a menu of three personalized goals (Goal 1, Goal 2, Goal 3) each associated with an all-or-nothing reward. Rewards are denominated in points that employees can later redeem for prizes in a marketplace. BIW promotes the program as having a 98 or 99 percent participation rate among eligible employees. While we cannot directly verify participation statistics, high participation rates are plausible due to marketing and communication during the pre-period, the administrative ease of enrollment, and the often-valuable rewards associated with goal attainment.

In 2014, we asked BIW to implement an enhanced enrollment process to elicit additional data from employees, including their beliefs about goal attainment. Under enhanced enrollment, respondents were prompted to complete a brief survey immediately after selecting their goal. The survey asked employees to estimate their perceived likelihood of attaining each of the goals: “On a scale from 0% (no chance) to 100% (absolute certainty), how likely is it that you will meet or exceed each of the following achievement levels?” (scale indexed in increments of 10 percentage points). Employees were additionally asked about their gender, age, and tenure with the firm. While the survey was optional, due to its integration within the enrollment process, survey participation across our sample was a robust 60 percent.

Following goal selection, employees transitioned to the several-week performance period during which they attempted to achieve their selected goal. In most programs, participants were able to log onto the website to check their progress or to remind themselves of their selected goal. At the close of the performance period, employees who attained their goal could exchange their reward points for an actual reward in the GQ marketplace. The non-monetary rewards included major electronics (e.g., a flat-screen television), event packages, vacations, household items (e.g., luggage), or recreational items (e.g., golf clubs). Employees were educated as to the approximate conversion rate between points and the dollar value of the associated rewards; for many programs, employees were familiar with the marketplace through other BIW programs using the same point currency.

2.2.3 Goal and reward structure

Several elements of the GQ goal and reward structure were designed to increase employee productivity, premised on research in behavioral science. First, GQ asked employees to self-select a goal from a personalized menu. Specifically, excepting employees without any experience, a personalized goal menu was generated by applying a uniform rule to an employee’s productivity during some baseline period prior to program administration (e.g., productivity during the prior quarter). Almost all program menus featured additively linear goals of the form: $f(x_b), f(x_b)+a, f(x_b)+2a$, where $f(x_b)$ is a function of baseline productivity, x_b, (e.g., $f(x_b) = 1.05x_b$) and a denotes some increment, potentially itself a function of baseline productivity (e.g., 10 or $0.10x_b$). To further increase personalization, employees within a program were typically segregated into a small number of distinct groups based on factors such as their baseline performance, experience, or job type. While goal menus within each group were personalized using a single rule, rules could vary across groups. For example, this segregation strategy permitted GQ to assign new employees to a menu not informed by baseline data or to use different rules for employees who differed widely in their baseline performance. To ensure goals were genuinely self-selected, employees were not nudged towards a particular goal during the goal-selection process via recommendations, defaults, or persuasion.

Second, based on the presumed motivational power of goals, GQ encouraged higher goal choice by designing them to be more financially attractive, in expectation, for most employees. In contrast to the typical menu with linearly increasing goals, rewards typically increased in non-linear increments. For example, many reward menus followed the k, $3k$, $6k$ structure, where k was set to be approximately 1 percent of an average employee’s salary over the course of the program. Moreover, goals were all-or-nothing such that an employee selecting Goal 3 and achieving the Goal 2, but not Goal 3, threshold would earn no reward while an employee selecting Goal 1 and achieving the Goal 3 threshold would only earn the Goal 1 reward. As a result of reward non-linearity, and the all-or-nothing design, we estimate that, under rational expectations, Goal 3 maximized expected value for 84 percent of employees (Goal 2 maximized expected value for 11 percent of employees). Finally, the rewards associated with each goal were non-monetary, due to a belief that non-monetary rewards were more motivating than monetary rewards.

2.3 Theoretical framework for goal choice

We now introduce a theoretical framework to organize our analysis of conservative goal choice. We represent the GoalQuest program as a choice between two simple lotteries and assume employees select the goal that maximizes their expected utility given their beliefs of goal attainment. We then amend the model to consider systematic departures from the standard framework such as the potential for biased beliefs, non-standard decision-weights, and reference-dependent utility. Finally, we consider the possibility that conservative choice can be attributed to employee heterogeneity in decision-making frameworks.

2.3.1 Generalized expected utility framework

We begin by outlining a generalized framework to describe how a utility-maximizing employee selects a productivity goal associated with an all-or-nothing reward from a menu of choices. To simplify matters, we model the menu as having just two options, a high goal (higher difficulty, higher reward) and a low goal (lower difficulty, lower reward). Employees understand that, in the period following goal selection they will earn the pre-specified reward if their level of productivity meets or exceeds their selected goal. We assume that the level of realized output is attributable in part to employee-specific ability, non-employee-specific productivity shocks, but does not depend on goal selection.

Formally, we represent goal choice as participation in one of two available lotteries, $G_n \in [G_h, G_l]$. Each lottery yields a reward x_n with some probability s_n and 0 with some probability $(1 - s_n)$. In our context, the high goal has a strictly higher reward and lower likelihood of attainment than the low goal, $x_h > x_l$ and $s_h < s_l$. Subjective probabilities, $\hat{s}_n$, capture employee beliefs about actual likelihoods of attainment. An employee must select exactly one of the two lotteries. We can describe an employee's valuation of a goal-lottery as follows:

$$V(G_n) = \pi_n(\hat{s}_n n)u(x_n, \theta) \tag{1}$$

Here, π_n denotes the decision weight an employee assigns to the subjective probability of goal attainment, and $u(.)$ is an always increasing function that captures an employee's preference rewards and potentially depends on a reference point θ.

An employee will choose the low goal under the following condition:

$$\pi_l v(x_l, .) \leq \pi_h v(x_h, .) \tag{2}$$

Our generalized framework identifies several potential reasons for why an employee might choose a conservative goal including a substantial difference in expected likelihood of attainment, the curvature of their utility function, biased beliefs about goal attainment ($\hat{s}_n \neq s_n$), and non-linear decision weights $\pi(\hat{s}_n) \neq \hat{s}_n$).

2.3.2 Conservatism with EU and rational expectations ($\pi(s) = s, v(x_n, .) = u(x_n)$)

As a baseline informed by expected utility theory, a well-informed employee selects the goal that maximizes expected utility using linear decision weights, a utility function dependent only on final wealth states, and rational expectations of attainment. For tractability, we assume a parametric utility function from the constant absolute risk aversion (CARA) family, so that the parameter, r, captures an employee's attitude towards risk, $r > 0$ implies risk aversion, and $r = 0$ denotes risk neutrality (we ignore the possibility that $r < 0$):

$$u(x_n) = \begin{cases} -\frac{1}{r}\exp(-r x_n), & r > 0 \\ x_n, & r = 0 \end{cases}$$

While our choice of a CARA function permits us to represent risk attitudes with a single parameter, it implies the irrelevance of an employee's prior wealth for risk preferences. We speculate that abstracting away from initial wealth is reasonable given that an employee must evaluate two lotteries relative to a single level of initial wealth. Our assumption of rational expectations implies that employees have unbiased and well-informed beliefs regarding the likelihood of goal attainment, $\hat{s}_n^r$, such that $\hat{s}_n^r = E[s_n|\Phi] = s_n + \varepsilon$. Here, Φ is the information set available to an employee at the time of goal choice and Φ is a normally distributed, mean-zero, error term with constant variance.

Risk neutrality ($r = 0$). For completeness, we first consider the case of risk neutrality. An employee who is indifferent to financial risk will choose the low goal if: $\hat{s}_l^r/\hat{s}_h^r > x_h/x_l$. Given these preferences and beliefs, we should expect to observe conservative goal choice only if the relative expected likelihood of achieving the low versus high goal exceeds the ratio of the high versus low goal.

Risk aversion ($r > 0$). Next we consider the more plausible scenario in which an employee is averse to financial risks. Such an employee will choose the low goal if:

$$r > \frac{\ln\left(\frac{\hat{s}_l^r}{\hat{s}_h^r}\right)}{x_l - x_h} \tag{3}$$

The decision rule implies that conservative goal choice is positively increasing in the degree of risk aversion, as well as expectations of relative goal attainment, and the gap between high and low goal rewards. We consider risk aversion parameters within some range of plausibility $r < r'$. Practically, we establish an upper bound of plausibility by examining the behavior implied by such risk preferences in simple lotteries involving financial stakes comparable to those engaged in the GQ program.

Risk aversion ($r_i \in [0, r']$). Finally, we consider the possibility that employees exhibit heterogeneity across their risk preferences. The possibility of non-uniform attitudes towards risk has been explored in prior research. We specifically consider whether the goal choices of employees can be rationalized by any degree of risk aversion within an interval of plausibility, $r_i \in [0, r']$.

2.3.3 Conservatism due to non-standard beliefs ($\hat{s}_n \neq E(s_n|\Phi)$)

We next consider the possibility that conservative goal choice emerges from the non-standard beliefs of a risk averse employee who maximizes an expected utility function with linear decision weights. We can depict non-standard beliefs with a multiplicative constant, $\hat{s}_n = \gamma_n s_n + \varepsilon$, such that γ_n represents the degree of goal-specific distortion to beliefs. As a result, $\gamma_n > 1$ implies overconfidence while $\gamma_n < 1$ implies underconfidence.

A risk averse employee with distorted choices of this nature will choose the low goal if:

$$r > \frac{\ln\left(\frac{\hat{s}_l^r}{\hat{s}_h^r}\right) + \ln\left(\frac{\gamma_l}{\gamma_h}\right)}{x_l - x_h} \tag{4}$$

The decision rule implies that conservative goal choice increases in γ_l/γ_h. If $\gamma_l/\gamma_h > 1$, then the share of conservative choice should be higher than that implied by the standard benchmark, while if $\gamma_l/\gamma_h < 1$, then the share of conservative choice should be lower than that implied by the standard benchmark. If distortion of beliefs is symmetric across the low and high goal, then the share of conservative choice should not differ from the benchmark.

2.3.4 Conservatism due to non-standard decision weights ($\pi(s) \neq s$)

We now assess whether the adoption of non-linear decision weights helps to explain employee behavior. In particular, researchers have advanced several probability weighting functions to address violations of expected utility in which people appear to overweight highly improbable outcomes and underweight highly probably outcomes. Given the literature's emphasis on an inverse-S shaped weighting functions, we adopt arguably the most popular of these functions, the function proposed by Prelec (1998):

$$\pi_n = \exp(-(-\ln s_n)^\alpha) \tag{5}$$

The non-linear weighting function could result in conservative goal choice if an employee were to underweight the likelihood of attaining a high goal relative to attaining a low goal. The decision rule for an employee governed by non-linear decision weights is given by:

$$r > \frac{\ln\left(\frac{\pi_l(s_l)}{\pi_h(s_h)}\right)}{x_l - x_h}. \tag{6}$$

2.3.5 Conservatism due to loss aversion ($u(.) = u(x_n, \theta)$)

Finally, we consider the possibility that conservative goal choices may arise as the result of employees exhibiting loss aversion in the context of gain-loss preferences (Kahneman and Tversky 1979; Tversky and Kahneman 1992)). Loss aversion has been advanced as a possible explanation for small- to moderate- scale risk aversion by Rabin and Thaler (2001) and has practically been suggested as an explanation for field evidence in contexts like insurance (Sydnor 2010), investments, and betting. Given that the structure of GQ stipulates that every employee receives either nothing or a positive reward, employees do not confront explicit losses in the program context. However, following the expectation-based approach of Kőszegi and Rabin (2006), it is reasonable to interpret goals, especially those associated with substantial rewards, as potential reference points (Heath, Larrick and Wu 1999). Specifically, given some reference point θ, we can describe an employee's utility by the following:

$$U(x_n) = \begin{cases} u^+(x_n - \theta), & x \geq \theta \\ -\lambda u^-(x_n - \theta), & x < \theta \end{cases}$$

As is standard practice, we specify that utility over gains, u^+, is concave, while utility over losses, u^- is convex. While the literature does not provide clarity as to how practically specify a reference point in the context of the goal menu, in the empirical analysis we consider a range of plausible reference points informed by both theoretical and exploratory work.

2.4 Data and sample construction

Our analysis of financial decisions under risk leverages program- and employee-level administrative data from BIW. The employee-level data included demographic detail, goal choice, employee productivity over the duration of the program, and employee beliefs elicited from enhanced enrollment. The program-level data identified each firm (and department), the date of program administration, rules used to segregate employees into groups, and details of the goal/reward menus faced by each participant. In this section, we describe the construction of our analytic sample, summarize its key features, and define the variables central to the subsequent analysis.

2.4.1 Primary employee sample

We constructed the primary employee sample by applying screening restrictions to an original dataset obtained from BIW. This original data, which spanned 38,661 employees across 34 programs and 18 firms, reflected the universe of data from GQ programs administered between 2014 to 2018 in the US or Canada that had adopted enhanced enrollment, had at least 100 participants who completed the program, and whose data had been electronically archived by BIW.[1] To generate the primary sample, we then restricted the original data to the 60 percent

[1] Due to a variety of circumstances, data for a small number of programs was not archived by BIW. The program size threshold was practically necessitated by the fixed administrative burdens to BIW of organizing and transferring program data as well as the significant resources required by our research team to process and manually audit data for each program.

of employees who participated in enhanced enrollment (and for whom, we therefore, had data on beliefs of goal attainment). We additionally excluded a modest number of records for which a data field was missing (excluding employee salary for which we only have partial coverage), we inferred the employee had likely not completed the program, or responses to the belief elicitations were internally inconsistent.[2]

We generated a secondary sample by amending the primary sample to include data without employee beliefs but that otherwise satisfied data restrictions. A comparison of demographics and goal choice across the primary and the secondary samples suggests no clear evidence for sample selection, at least based on observable factors.

The primary sample comprises the decisions of 20,133 employees across 18 firms, 34 programs, and 232 distinct groups. Table 2.1 summarizes overall sample statistics as well as group-level (duration, financial stakes) and employee-level (age, gender, tenure, inferred income) characteristics. On average, we observe data for 1,158 employee participants per program (IQR: 208 to 703) and 308 employees (IQR: 12 to 103) per group. Groups varied roughly uniformly across three categories of duration, 30, 60, and 90 days (there are two outlier programs that ran for 45 and 120 days). While the group-level average (median) potential reward was $607 ($375), the distribution of potential reward values was asymmetric, such that 10 percent of employees faced decisions involving rewards worth an average of $2,150. Overall, employees in the sample could have earned up to $9.4 million in possible rewards.

The table also indicates the diversity of the employee sample with respect to gender, age, and tenure. We speculate that the high share of employees, 72 percent, with tenure of 5 years or less likely reflects the fact that GQ programs are often administered in industry sectors with high turnover. While we only observe average program-level salaries for programs in 8 of the 18 firms, or 25 percent of the sample, the average of this program-level statistic is $70,400, placing employees at the 72nd percentile of the U.S. national individual income distribution as of 2018. However, because the availability of salary data is greater for programs with higher rewards, we suspect the average salary among employees in our sample is closer to the national average for full-time employees.

2.4.2 Goal choice and employee productivity

We turn next to data on goal choice and employee productivity. We describe goal choice, g, both through indicators for goal choice as well as indicators denoting whether the goal reflects an optimal, aggressive, or conservative choice. The characterization of goal choice relies on a comparison of the expected utility associated with the selected, and non-selected, goals with respect to the benchmark models outlined in the theoretical framework. We describe employee productivity through two normalized measures: (i) productivity relative to the baseline threshold and (ii) productivity relative to the Goal 3 threshold. Normalization facilitates comparisons across programs whose productivity outcomes may vary substantially in their scale (e.g., at a call center, productivity may be measured in hundreds of calls over the program period, while at a sales office, it may be measured in single-digit unit sales). Finally, we report indicators of baseline and goal attainment for each employee.

Table 2.2 summarizes the choice, productivity, and attainment measures. The table indicates that 44 percent of employees selected the highest goal while remaining employees roughly split across a choice of the low and medium goals. The performance of the median employee nominally exceeded the baseline threshold (1.01 ratio) and moderately fell short of the Goal 3 threshold (0.89 ratio). Said differently, while just over one-half of employees reached their baseline threshold, only 29 percent reached the Goal 3 threshold. The table also indicates a correlation between goal choice and productivity – for example, 40 percent of those selecting Goal 3 attained the goal – suggesting either that higher goals led to an increase in performance or that employees with higher expected productively positively selected into higher goals.

[2] A total of 508 respondents (2.6%) were eliminated for these reasons. An employee was excluded for inconsistent beliefs if she responded that she believed she was strictly more likely to attain a higher goal than a lower goal.

Table 2.1 · Summary of sample, group, and employee characteristics

		Potential reward value	
	All	Below median	Above median
Panel A. Sample overview			
Programs	34	-	-
Groups	232	-	-
Employees	20133	-	-
Firms	18	-	-
Employees per group (average)	308	-	-
	(277.3)		
Employees per grogram (average)	1158	-	-
	(947.7)		
Panel B. Group characteristics (employee shares)			
Program duration			
$\leq$ 30 days	0.39	0.51	0.28
45 to 60 days	0.28	0.12	0.42
$\geq$ 90 days	0.33	0.38	0.29
Potential reward value (estimated $)			
Average	466.5	150.1	746
	(481.5)	(57.8)	(516.8)
Median	350	168	525
25th percentile	175	94	392
75th percentile	525	175	914
Panel C. Employee characteristics			
Age (midpoint of 10-year bins)	36.9	36	37.6
Female	0.46	0.50	0.43
Tenure category			
< 1 year	0.28	0.32	0.25
1 to 5 years	0.45	0.46	0.43
6 to 10 years	0.14	0.13	0.14
> 10 years	0.13	0.08	0.18
Program-average salary (average) ($1,000s)	70.8	63.2	72.7
Data on salary available	0.25	0.10	0.38

This table summarizes observable detail on GQ programs and employees. Specifically, Panel A describes the number and size of programs across the overall analytic sample, while Panel B describes average program duration and average potential reward values at the employee-level. Potential reward value refers to the largest reward an employee can potentially earn in the program, or alternatively, the value of the Goal 3 reward. Panel C summarizes demographic details of employees including age, gender, tenure, and approximate salary for all employees and by sub-groups distinguished by potential reward value. We impute age from self-reported 10-year bins, infer gender using a combination of self-reported data and inference from first name, and approximate salary using program-level averages for those programs for which data was available.

Table 2.2 · Goal choice, employee productivity, and goal attainment

		Sample restricted by goal choice		
	All	**Goal 1**	**Goal 2**	**Goal 3**
Panel A. Goal choice				
Employees	20133	5866	5470	8797
Employee share	1.00	0.29	0.27	0.44
Potential reward value (average), $	466	482	490	442
	(481.5)	(528)	(499)	(434.4)
Panel B. Employee productivity				
Productivity relative to baseline				
Average	1.34	1.12	1.25	1.52
25th percentile	0.88	0.78	0.89	0.91
50th percentile	1.01	0.98	1.00	1.04
75th percentile	1.20	1.11	1.15	1.27
Productivity relative to Goal 3 threshold				
Average	0.90	0.68	0.86	1.07
25th percentile	0.60	0.30	0.63	0.77
50th percentile	0.89	0.74	0.88	0.95
75th percentile	1.02	0.95	1.00	1.09
Panel C. Goal attainment				
Baseline	0.54	0.45	0.53	0.60
Goal 1	0.44	0.32	0.42	0.53
Goal 2	0.36	0.23	0.33	0.47
Goal 3	0.29	0.17	0.25	0.41
Earned reward (average)	121	33	92	197
Earned reward (average) | goal attainment	333	104	277	483

This table summarizes goal choice, productivity, and goal attainment for the overall sample and separately by employee goal choice. Specifically, Panel A summarizes goal choice and average potential rewards, Panel B summarizes employee productivity relative to baseline and to Goal 3, and Panel C summarizes goal attainment and average earned rewards. Potential reward value refers to the largest reward an employee could potentially earn in the program, or alternatively, the value of the Goal 3 reward. The summary of productivity relative to baseline excludes the 18 percent of employees without baseline data.

2.4.3 Employee beliefs

Our analysis draws on two measures of employee beliefs: raw subjective beliefs elicited through enhanced enrollment and estimates of ex ante rational expectations. To construct a measure of subjective expectations describing employee i's probabilistic belief, $\hat{s}_{(k,i)}$, of attaining goal threshold k, we simply record the employee's response in enhanced enrollment, in 10-unit increments on a scale from 0 to 100 percent. Our estimation of an employee's ex ante rational expectation of goal attainment, $\hat{s}^r_{(k,i)}$, consists of two steps. First, we calculate the average ex post goal attainment for each program group and each goal. Next we predict the ex ante likelihood of goal attainment for each goal and each employee by adjusting the group-average by observable covariates. The exercise effectively assumes that one can proxy for rational expectations by appealing to average attainment for similar others. The estimation strategy is consistent with the intent of program administrators to sort employees with similar performance expectations (or, in the case of new employees, a similar lack of baseline data) in the same group.

To implement the estimation strategy we first estimate the following leave-out regression for each employee

Table 2.3 · Employee beliefs and confidence of goal attainment

		By goal choice		
	All	**Goal 1**	**Goal 2**	**Goal 3**
Panel A. Employee beliefs				
Rational expectations of goal attainment				
Goal 1	0.44	0.41	0.44	0.46
Goal 2	0.37	0.32	0.36	0.39
Goal 3	0.30	0.25	0.28	0.33
Subjective beliefs of goal attainment				
Goal 1	0.78	0.65	0.79	0.86
Goal 2	0.69	0.50	0.71	0.82
Goal 3	0.63	0.43	0.57	0.77
Panel B. Over/under confidence				
Ratio of subjective/rational beliefs				
Goal 1	2.20	2.09	2.26	2.27
Goal 2	2.62	2.42	2.79	2.76
Goal 3	3.46	3.26	3.43	3.59
Relative ratio of over/under confidence				
Goal 3/Goal 1	1.45	1.41	1.42	1.48
Goal 3/Goal 2	1.22	1.24	1.15	1.22
Goal 2/Goal 1	1.13	1.08	1.18	1.17

This table summarizes employee beliefs and confidence with respect to goal attainment for the overall sample and separately by employee goal choice. Specifically, Panel A successively summarizes beliefs of goal attainment under rational expectations and then under subjective beliefs. We assign employee- and goal-specific rational expectations by adjusting the ex post average rate of goal attainment at the group-level by employee age and gender, as estimated from a linear regression (for a small share of employees, for whom this strategy violated monotonicity, we adopted the unadjusted ex post average). Subjective beliefs for each goal reflect employee self-reports, elicited during enhanced enrollment, using an eleven-point scale (0, 10, 20, . . . , 100 percent). For tractability, we adjust any belief of 0 or 100 percent to 1 and 99 percent, respectively. Panel B summarizes employee under/over confidence as conveyed by the average ratio of subjective beliefs and rational expectations, such that a ratio > 1 indicates overconfidence. To minimize the effects of outliers, we Winsorized ratios by capping outliers below the 5th percentile and above the 95th percentile. Finally, the panel reports relative under/over confidence across specific goal pairs, as conveyed by the average ratio of Winsorized under/over confidence.

i and goal threshold $k \in [0, 1, 2, 3]$:

$$I[s_{k,l,-i}] = \alpha + Z\gamma + \pi_l + \epsilon \tag{7}$$

Each regression predicts average group-level attainment for each goal, $s_{k,l,-i}$), leaving out employee i, as a function of employee characteristics included in vector Z (age, tenure, gender) and group fixed effects, π_l. We then calculate an employee's rational expectation of obtaining goal k, as $\hat{s}^r_{k,-i} = \hat{\alpha} + Z\hat{\gamma} + \hat{\pi}_l$. We estimate regressions at the program level so that we have a larger sample from which to estimate employee covariates. This technique of inferring rational expectations by looking at the realized outcomes of similar others is common in economic analyses of insurance choice (e.g. Bhargava, Loewenstein and Sydnor 2017).

Table 2.3, which summarizes belief data, indicates two patterns of interest. First, the table suggests that increasing expectations of attainment predicted goal choice. Employees selecting more aggressive goals had, on average, higher (rational and subjective) expectations of attaining those goals. For example, 43 percent of those selecting Goal 1 subjectively expected to attain Goal 3, compared to 76 percent of those selecting Goal 3. Second, the comparison between subjective beliefs and our estimates of rational expectations suggests that employees were substantially overconfident, on average, about their future productivity. We investigate employee overconfidence in greater depth in the next section.

2.5 Characterization of goal choice by benchmark model

We now turn to characterizing employee choice using the decision-making benchmarks outlined in the theoretical framework. For each benchmark model, we report the share of conservative, optimal, and, for completeness, aggressive, goal choice. To better understand the moderating role of economic stakes and experience on goal choice, we additionally report the share of optimal choice across potential reward size and employee tenure.

2.5.1 Expected utility with risk neutrality

We begin by assessing choice relative to a baseline benchmark which presumes that risk-neutral employees choose a goal that maximizes their expected utility given beliefs of goal attainment. We initially assume perfect information before considering the case of rational expectations and subjective beliefs, directly elicited via enhanced enrollment. For each benchmark, we report the share of goal choices that match the prediction of the model (optimal choice), the share of goal choices that involve a goal lower than the model prediction (conservative choice), and the share of goal choices that involve a goal higher than that predicted by the model (aggressive choice).

Perfect information benchmark. We first consider the scenario in which employees have perfect information regarding their eventual productivity. The unrealistic, but instructive, exercise is equivalent to characterizing choice based on an employee's observed ex post productivity, assuming that productivity is unaffected by goal choice. We restrict this analysis to the 44 percent of employees who attained Goal 1 since, for remaining employees, the ex post benchmark implies the irrelevancy of choice. The first panel of Table 2.4 characterizes goal choice for the baseline benchmark model across different assumptions regarding beliefs. The table indicates that, assuming perfect information, just over one-half of employees chose optimally, while 31 percent of employees chose conservatively. For employees who attained Goal 1, sub-optimal choice resulted in an average unrealized reward of $272, where we define unrealized as the difference between the realized reward and the counterfactual reward one would have achieved had they chosen optimally. Figure 2.2, which depicts the cumulative distribution of unrealized rewards by goal choice, for employees achieving at least Goal 1, indicates that most unrealized rewards are concentrated among employees who selected either Goal 1 or Goal 2. In this sense, the figure confirms the optimality, under perfect information, of Goal 3 for most employees. Finally, to better understand the role of economic moderators, the second panel of the table reports how the optimality of choice varies across the potential reward stakes (indexed by quartile) and employee tenure (indexed by response category). The table implies that the share of optimal choice did not substantially increase with higher potential rewards or increased employee experience.

Rational expectations. Table 2.4 characterizes goal choice for the baseline benchmark model across different assumptions regarding beliefs. The table indicates that, assuming perfect information, just over one-half of employees chose optimally, while 31 percent of employees chose conservatively. For employees who attained Goal 1, sub-optimal choice resulted in an average unrealized reward of $272, where we define unrealized as the difference between the realized reward and the counterfactual reward one would have achieved had they chosen optimally. Figure 2.2, which depicts the cumulative distribution of unrealized rewards by goal choice, for employees achieving at least Goal 1, indicates that most unrealized rewards are concentrated among employees who selected either Goal 1 or Goal 2. In this sense, the figure confirms the optimality, under perfect information, of Goal 3 for most employees. Finally, to better understand the role of economic moderators, the second panel of the table reports how the optimality of choice varies across the potential reward stakes (indexed by quartile) and employee tenure (indexed by response category). The table implies that the share of optimal choice did not substantially increase with higher potential rewards or increased employee experience.

Subjective beliefs. We proceed to consider the possibility that the high share of apparent conservatism implied by the rational expectation benchmark may reflect systematic bias in employee forecasts of productivity. Such bias could account for conservative choice if employees held inflated beliefs about their relative likelihood of attaining

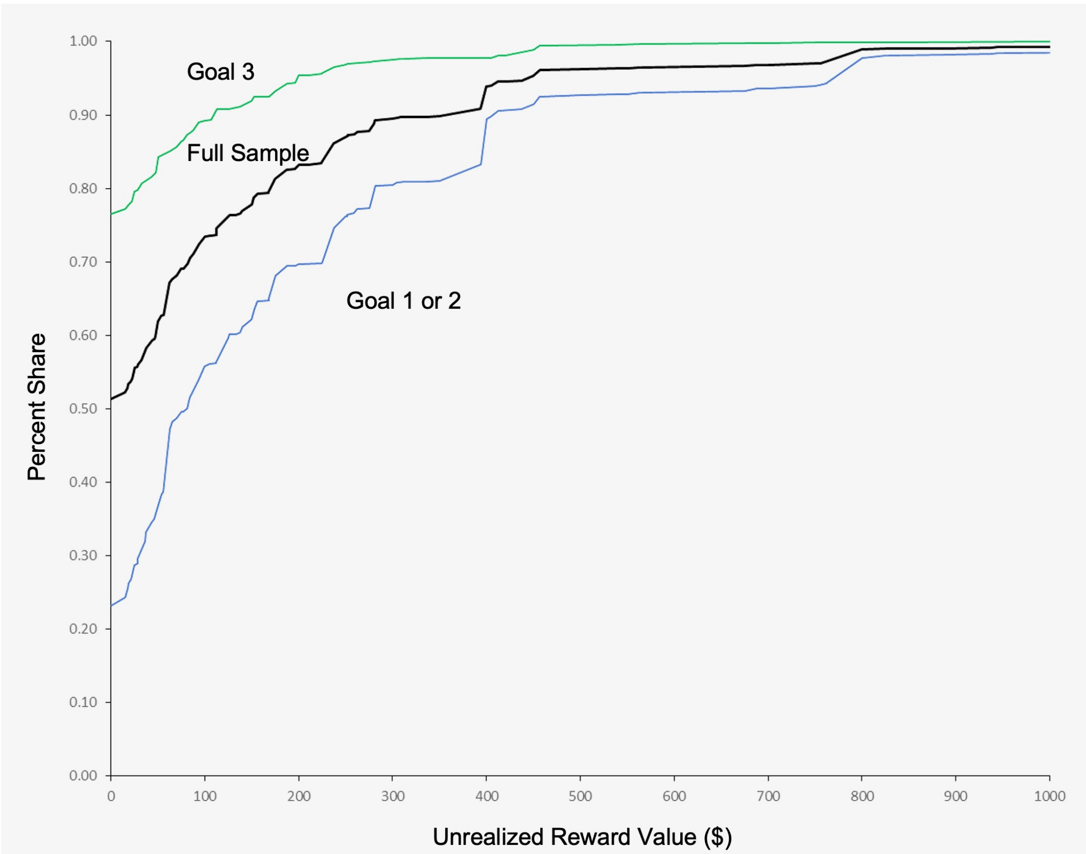
Table 2.4 · Goal choice characterization for expected utility benchmarks

				Absolute Risk Aversion (CARA)					
	Risk Neutral			Rational expectations			Subjective beliefs		
	Perfect	Rational	Subjective	$r = 0.0003$	$r = 0.005$	$r < 0.005$	$r = 0.0003$	$r = 0.005$	$r < 0.005$
Panel A. Characterizing goal choice									
Optimal choice	0.51	0.45	0.50	0.45	0.44	0.55	0.50	0.53	0.59
Conservative choice	0.31	0.49	0.48	0.49	0.38	–	0.48	0.39	–
Aggressive choice	0.18	0.06	0.02	0.06	0.17	–	0.02	0.08	–
Expected reward | chosen goal	274	109	214	109	109	–	214	214	–
Expected reward | optimal choice	381	146	275	146	124	–	275	247	–
Maximum expected loss	2813	1435	2272	1435	998	–	2272	2272	–
Unrealized reward given sub-optimal choice + goal attainment	272	139	120	140	122	–	118	104	–
Panel B. Optimal choice share									
Potential reward value									
Highest quartile	0.53	0.42	0.48	0.42	0.39	0.62	0.49	0.55	0.69
Lowest quartile	0.48	0.44	0.48	0.44	0.44	0.46	0.48	0.48	0.49
Employee tenure									
Highest category (10+ years)	0.49	0.39	0.45	0.40	0.40	0.51	0.46	0.53	0.60
Highest category (1 years)	0.52	0.44	0.47	0.44	0.44	0.58	0.47	0.50	0.62

Notes: This table characterizes the efficiency of employee goal choice under expected utility across a range of assumptions regarding CARA risk preferences and employee beliefs (perfect, rational, and subjective). Specifically, Panel A characterizes employee choices as either optimal, conservative, or aggressive relative to the prediction of the benchmark model and additionally reports average expected and unrealized rewards conditional on choice. Panel B reports the share of optimal choice across employee sub-groups distinguished by the size of the potential reward and years of experience. Characterization of choice under the perfect information benchmark excludes employees who did not attain Goal 1 (characterization under other benchmark models rely on the entire sample). The blank cells reflect the inability to uniquely characterize aggressive and conservative choices for benchmarks involving flexible values of r.

Figure 2.2 · Cumulative distribution of unrealized rewards: ex-post attainment

This figure depicts the cumulative distribution of unrealized rewards overall and separately by goal choice for the 8,800 employees whose productivity met or exceeded the Goal 1 threshold. An unrealized reward refers to the difference between an employee's actual earned reward and the counterfactual reward an employee could have earned if they had chosen ex post optimally. By definition unrealized rewards cannot be negative. While the figure censors unrealized rewards at $1,000 for clarity, a small share of employees had unrealized rewards in excess of $1,000, with a maximum unrealized reward of approximately $2,800.

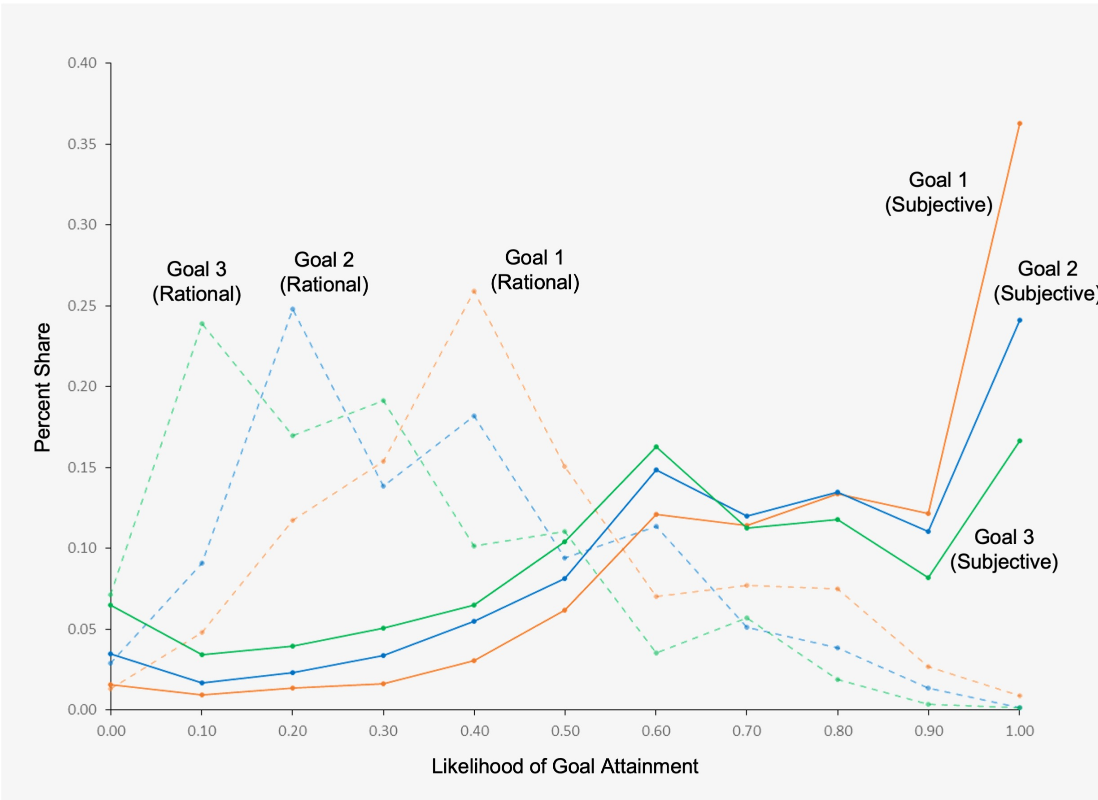


low versus high goals. For example, if employees were systematically under-confident about future productivity, and this led employees to inflate the likelihood of achieving lower, relative to higher goals, then one might expect utility-maximizing employees to select lower goals more frequently than predicted by the benchmark model. Alternatively, employee overconfidence could similarly explain the high share of ostensibly conservative choice if such overconfidence led to employees to systematically inflate the likelihood of achieving lower, relative to higher, goals. To assess whether the observed choice patterns reflect biased beliefs, we re-characterized goal choice after replacing rational expectations of goal attainment with subjective beliefs elicited from the enhanced enrollment process. The results, reported in Table 2.4, indicate that adopting subjective, rather than rational, expectations, did not reduce the share of conservative choice and led to only modest improvement to overall descriptive accuracy. As with the prior benchmarks, reward size and employee experience did not predict more efficient choice in the subjective utility benchmark.

Table 2.3 provides additional insight into how employee beliefs affect the characterization of choice. For each goal, the table reports the average rational and subjective beliefs of attainment and the average ratio of subjective and rational expectations for the entire sample and separately by goal choice. We highlight three patterns of

Figure 2.3 · Distribution of rational expectations and subj. beliefs of goal attainment

Notes: This figure compares the distributions of rational expectations and subjective beliefs of goal attainment for each goal. We assign employee- and goal-specific rational expectations by adjusting the ex post average rate of goal attainment at the group-level by employee age and gender, as estimated from a linear regression (for a small share of employees, for whom this strategy violated monotonicity, we adopted the unadjusted ex post average). Subjective beliefs for each goal reflect employee self-reports, elicited during enhanced enrollment, using an eleven-point scale (0, 10, 20, . . . , 100 percent). For ease of comparison, the figure groups rational expectations into bins that parallel the subjective belief data.



note from the table. First, Panel A indicates that subjective beliefs of goal attainment strongly predicts goal choice, suggesting that the elicitation produced credible results. Second, Panel B indicates that subjective beliefs of attainment substantially exceed rational expectations for each of the three goals by a ratio ranging from 2.20 (Goal 1) to 3.46 (Goal 3). This suggests that employees exhibit substantial overconfidence across all of the goals. Figure 2.3 corroborates the high degree of overconfidence across each goal.

Finally, the ratios of relative overconfidence reported in the second panel of the table suggest that employees are, on average, more overconfident about attaining higher, relative to lower, goals. Ultimately, the table and figure make clear that one cannot explain the high share of conservative goal choice from the expected utility framework through biased employee beliefs – employees are absolutely, and relatively, overconfident about attaining high goals.

2.5.2 Expected utility with risk aversion

We turn now to the possibility that conservative choice may reflect risk aversion attributable to the diminishing marginal utility of wealth. As described in the theoretical framework, we model risk aversion by assuming parametric utility that exhibits constant absolute risk aversion, the degree of which is indicated by the parameter

r. We represent the plausible range of risk aversion with the interval $r \in [0.0003, 0.005]$. To appreciate the breadth of risk attitudes captured by this interval, one can translate what such risk preferences imply for gambles involving potential losses of a size similar to the potential rewards engaged by employees in GQ. For example, consider a simple lottery involving a 50 percent chance of losing \$175 (the 25th percentile GQ reward value) and a 50 percent chance of some unspecified gain. A risk aversion parameter of $r = 0.0003$ implies that an employee would accept any such gamble so long as the potential gain exceeds \$184 – a modest, but seemingly plausible, degree of risk aversion. The same employee should accept any 50/50 gamble involving a potential loss of \$350 (the median GQ reward value) so long as the potential gain exceeds \$391. At the other endpoint, $r = 0.005$ implies that an employee would reject any 50/50 gamble involving a potential loss of \$175 (or \$350), even if the potential gain was infinite. We therefore treat $r = 0.005$ as an upper bound of plausible risk aversion for financial stakes in the range of interest.

Table 2.4 characterizes choice for risk-averse, expected utility maximizing, employees under either rational or subjective expectations. Perhaps unsurprisingly, the assumption of modest risk aversion (r = 0.0003) does little to shift the characterization of choice, across either information regime, relative to risk neutrality. However, across both information regimes, the assumption of severe risk aversion (r = 0.005) shifts the characterization by moderately reducing the apparent share of conservative choice, largely offset by an increase in the share of choice characterized as aggressive. Incorporating risk aversion into the benchmark model does not substantially shift the descriptive accuracy of the model (i.e., the share of optimal choice). Risk aversion also does not affect the continued absence of moderation by reward size or employee experience.

Figure 2.4 conveys the intuition for the shift in choice characterization under risk-averse preferences. The figure depicts the share of employees whose goal choice is deemed as optimal by the benchmark model overall, and by goal choice, assuming rational (Panel A) and subjective (Panel B) beliefs, for r ranging from $r = 0.0$ to $r = 0.10$. Across panels, the figure shows that as one increases the degree of assumed risk aversion within the plausible range (highlighted region), the share of optimal choice among employees choosing Goals 1 and 2 increases moderately, but this increase is offset by the reduced share of optimality among employees choosing Goal 3. Seen another way, greater risk aversion reduces the share of conservative choice among low-goal choosers and increases the share of aggressive choice among Goal 3 choosers. Ultimately, the exhibits suggest that the incorporation of plausible risk aversion, induced from the concavity of the utility function, does not improve the descriptive fit of the benchmark models and does not explain the conservatism of nearly 40 percent of employees.

Heterogeneous risk preferences. Conceivably, goal choice may reflect the utility-maximizing behavior of a population with heterogeneous risk preferences. To evaluate this possibility, we reassessed the optimality of choice after classifying any goal choice as optimal if it could be rationalized by any r within the interval, $r = [0, 0.005]$. As reported in Table 2.4, shifting from a benchmark model assuming extreme, but uniform, risk aversion, to one assuming heterogeneous risk preferences increases the share of optimal goal choice from 0.44 to 0.55 percent (rational expectations) and from 0.53 to 0.59 percent (subjective beliefs). Allowing for highly flexible risk preferences also serves to increase the differential share of optimal choice across high and low reward size but not high and low employee experience. We note that the differential rate of optimal choice across reward size may be mechanical, since for menus with higher rewards, the likelihood that flexible risk preferences rationalizes the choice of two of the three goals on the menu is higher than it is for menus with smaller rewards.

2.5.3 Non-linear decision weights

We proceed to consider whether two commonly invoked behavioral departures from the standard expected utility framework can help to explain employee choice. The first departure we consider is the assumption of non-linear decision weights (i.e., a model of rank-dependent utility). Specifically, we consider a weighting function suggested by Prelec 1998 ($\alpha = \beta = 0.65$) with an inverse s-shaped form commonly asserted in the literature. In theory, if employees systematically underweight moderate-probability outcomes, such as the attainment of Goal 3, relative to higher-probability outcomes, such as the attainment of Goals 1 and 2, then the assumption of a non-linear weighting function might help to better explain goal choice.

Figure 2.4 · Optimal Choice under EU by risk preference and information regime

This figure depicts the share of optimal choice overall and separately by goal choice under expected utility across varying levels of the CARA risk aversion parameter, r, and information regimes. Specifically, Panel A depicts the share of optimal choice assuming rational expectations for an extended range of r on a logarithmic scale while Panel B depicts the analogous characterization of choice under the assumption of subjective beliefs. The shaded region denotes the range of substantial but still plausible risk aversion, $r \in [0.0003, 0.05]$.

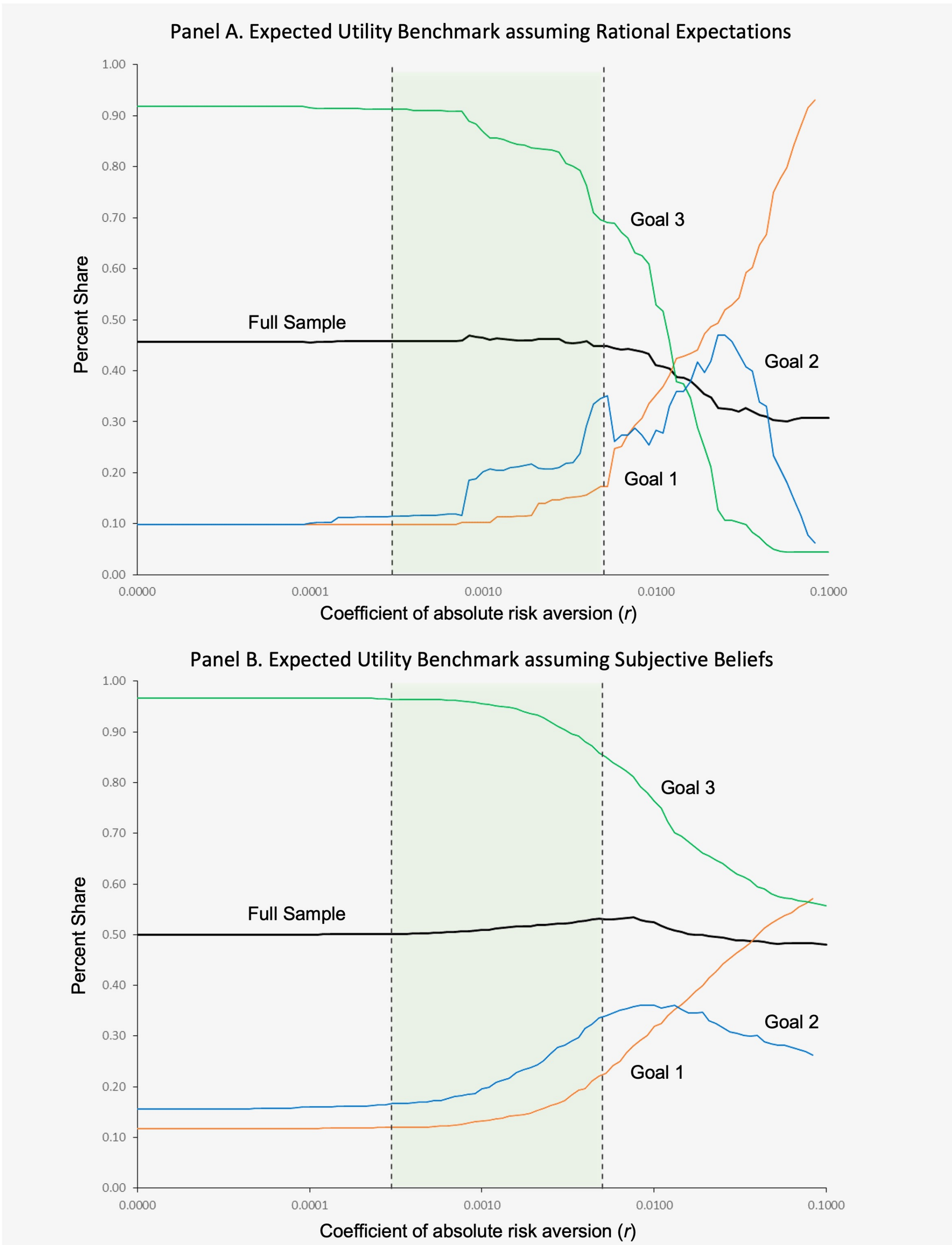


Table 2.5 contrasts the characterization of choice assuming non-linear decision weights relative to a baseline model of subjective expected utility with modest CARA risk aversion, $r = 0.0003$ (reported in the first column of Table 2.5 and reproduced from Table 2.4). The exercise indicates that non-linear decision weights do not meaningfully shift the characterization of choice relative to the subjective utility baseline. Given that few employees report extreme perceived probabilities of goal attainment – the segments of the probability distribution where decision weights depart most strongly from linear weighting – the absence of a strong shift in characterization with the adoption of non-linear weights is not surprising. Non-linear decision weights also do not change the previously documented absence of moderation across reward size or employee experience.

2.5.4 Gain-loss utility

Finally, we consider the possibility that conservative employee choice may reflect loss aversion in the context of gain-loss utility. While our context involves no explicit losses, research has suggested that goals may serve as reference points that invoke behavioral response consistent with those described in gain-loss paradigms (Heath, Larrick and Wu 1999). One practical challenge for assessing models of gain-loss utility, however, is the absence of clear theoretical guidance as to the functional implementation, magnitude of the loss aversion parameter, and the specification of the reference point. Regarding the latter, while Kahneman and Tversky (1979) originally adopted the standard quo as a reference point, they contemplated the potential for other reference points. Subsequent work has suggested a range of candidate reference points including those that are prospect-specific, expectation-based (Kőszegi and Rabin 2006; Loomes and Sugden 1986), and/or informed by salient considerations such as the certainty equivalence of a gamble (Gul 1991) or features of the choice menu. Perhaps a more practical resource for identifying reference points is provided by Baillon, Bleichrodt and Spinu (2020) who assess the success of gain-loss utility models across several prospect-independent (e.g., status quo, the high outcome, the highest probability option, the highest option a person is certain to achieve) and prospect-dependent (e.g., the selected option, the expected value of the selected option) reference points in explaining the risky choices of experimental subjects from a menu-based setting.

In light of our broad interest in all credible formulations of gain-loss utility, we aspired to characterize goal choice for a wide range of plausible reference points informed by the theoretical and empirical literature as well as the configuration of the GQ menu. We consequently considered five prospect-independent reference points: status quo (i.e., \$0), the high probability goal (Goal 1), the high reward goal (Goal 3), the highest goal an employee felt certain to achieve (\$0 for employees uncertain of achieving any goal), and, for completeness, Goal 2. We also considered prospect-dependent reference points including the chosen goal, the expected value of the chosen goal, and given the possibility that goals may emerge from pair-wise comparisons of contiguous goals, the proximal goal either below or above the chosen goal.

We assess these reference points in the context of two prominent functional representations of gain-loss utility in isolation, using the Kahneman and Tversky (1979) power function (initially setting $\alpha = 0.88$) and a composite framework in which utility is comprised of both consumption utility and a gain-loss component (Sugden 2003; Köbberling and Wakker 2005; Kőszegi and Rabin 2006, 2007). For the composite functions, we assume that consumption utility and gain-loss utility are additively separable, adopt the same KT power function for both utility components, and denote η as a scaling factor applied to consumption utility such that setting $\eta = 0$ reduces to a model with gain-loss utility only. In deference to the range of loss aversion parameters contemplated by the literature, and the lack of clarity as to the appropriate weighting across the two components of utility, we each model for values of λ from 1.5 to 3.0 and η from 1 to 5. Finally, to streamline the analysis, we ignore probability-weights across all of the gain-loss models and evaluate each model assuming subjective expectations.

Table 2.6 assesses the descriptive accuracy of the candidate formulations involving gain-loss utility. Among the prospect-independent reference points, nearly all of the reference points explain approximately one-half of all goal choices. To understand the intuition for this consistency, consider that for many employees, adopting a reference point shifts the utility-maximizing goal from Goal 3 to either Goal 1 or Goal 2. The net result is to increase optimality among low-goal choosers by roughly the same magnitude as the decrease in optimality

Table 2.5 · Goal choice characterization for non-standard benchmarks

	Baseline subjective EU [$r = 0.0003$]	Rank-dependent expected utility [Prelec $\alpha = \beta = 0.65$]	Gain-Loss Utility [$\alpha = 0.88, \lambda = 2.25$]		
			Gain-loss only [RP = g+1; η = 0]	Composite utility [RP = g; η = 1]	Counterfactual regret [RP CF; η = 0]
Panel A. Characterizing goal choice					
Optimal Choice	0.50	0.47	0.55	0.59	0.50
Conservative Choice	0.48	0.52	0.23	0.24	0.48
Aggressive Choice	0.02	0.01	0.22	0.17	0.02
Expected Reward | Chosen Goal	214	188	214	214	214
Expected Reward | Optimal Choice	275	244	230	244	275
Maximum Expected Loss	2272	2135	2272	2272	2272
Unrealized Reward given sub-optimal choice + goal Attainment	118	120	126	128	120
Panel B. Optimal Choice Share					
Potential Reward Value					
Highest Quartile	0.49	0.44	0.54	0.61	0.48
Lowest Quartile	0.48	0.46	0.53	0.55	0.48
Employee tenure					
Highest Category [10+ Years]	0.46	0.42	0.53	0.59	0.46
Lowest Category [# Years]	0.47	0.44	0.55	0.59	0.47

This table characterizes the efficiency of employee goal choice under a range of non-standard benchmark models. The first column of the table reproduces a baseline characterization from the expected utility benchmark assuming moderate CARA risk aversion ($r = 0.0003$) and subjective beliefs. The second column characterizes choice assuming a rank-dependent utility function, which modifies the baseline benchmark by applying nonlinear weights to potential outcomes using the Prelec function (Prelec 1998). The final three columns characterize choice under exemplar benchmarks with gain-loss utility (but not nonlinear weighting) using the Kahneman and Tversky (1979) value function (see text for details). Specifically, Panel A characterizes employee choices as either optimal, conservative, or aggressive relative to the prediction of the benchmark model and additionally reports average expected and unrealized rewards conditional on choice. Panel B reports the share of optimal choice across employee sub-groups distinguished by the size of the potential reward and years of experience. Characterization of choice under the perfect information benchmark excludes employees who did not attain Goal 1 (characterization under other benchmark models rely on the entire sample).

Table 2.6 · Descriptive accuracy of gain-loss utility models by candidate reference point

	Gain-loss utility ($\alpha = 0.88$)			**Consumption + G/L utility** ($\alpha = 0.88, \lambda = 2.25$)			
	λ = 1.50	λ = 2.25	λ = 3.00	η = 1	η = 2	η = 3	η = 5
Panel A. Prospect-independent reference points							
Status quo (0)	0.50	0.50	0.50	0.50	0.50	0.50	0.50
Min-max reward (Goal 1)	0.52	0.54	0.55	0.51	0.51	0.50	0.50
Compromise goal (Goal 2)	0.50	0.52	0.52	0.50	0.50	0.50	0.50
Max reward (Goal 3)	0.49	0.49	0.49	0.49	0.49	0.49	0.49
Max-min reward	0.51	0.51	0.51	0.50	0.50	0.50	0.50
Any prospect-independent reference point	0.53	0.57	0.61	0.53	0.51	0.51	0.51
Panel B. Prospect-dependent reference points							
Reward of chosen goal	0.29	0.29	0.29	0.59	0.57	0.56	0.54
Expected value of chosen goal	0.40	0.26	0.26	0.54	0.51	0.50	0.50
Reward of chosen goal + 1	0.55	0.55	0.55	0.53	0.53	0.53	0.52
Reward of chosen goal - 1	0.46	0.43	0.42	0.58	0.56	0.54	0.53
Regret (max exp. counterfactual)	0.50	0.50	0.50	0.50	0.50	0.50	0.50
Any prospect-dependent *reference point*	0.82	0.80	0.80	0.72	0.63	0.60	0.57
Any model	0.84	0.84	0.85	0.73	0.64	0.60	0.57

This table assesses the descriptive accuracy of benchmark models involving gain-loss utility across several candidate reference points, functional forms, and parameter specifications. The first set of columns characterizes choice under benchmark models involving gain-loss utility following Kahneman and Tversky (1979) across potential values of the loss aversion parameter, λ. The second set of columns characterizes choice under benchmark models involving composite utility, an additively linear combination of consumption utility and gain-loss utility, across potential consumption utility scaling factors, n ($n = 0$ therefore implies a model with gain-loss utility only). All benchmark models assume subjective beliefs. Panel A reports the share of optimal choice for prospect-independent candidate reference points while Panel B reports the analogous share of optimal choice for prospect-dependent candidate reference points. Please see text for additional detail on each of the benchmark models. .

among employees selecting goal 3. Among prospect-dependent reference points, the chosen goal provides the most successful reference point with respect to explaining goal choice. Across the composite utility formulations, the reference point explains between 54 and 59 percent of goal choices. The table also assesses the possibility that employees may vary in their strategy for forming reference points. This degree of flexibility increases the explanatory power of the benchmark considerably, particularly when one allows for any prospect-dependent model. However, once again, this effect may be somewhat mechanical since for many employees, such flexibility accommodates 2 or even all 3 goal choices.

Table 2.5 characterizes choice for the most promising of the gain-loss only and composite formulations. In deference to the prominence of models of regret in the literature, the table also reports the characterization associated with the counterfactual regret model. The table indicates that the incorporation of loss aversion via gain-loss utility delivers, at most, a modest increase in explanatory power relative to baseline. Once again, the table indicates that neither reward size nor employee experience moderate the descriptive accuracy of the models. Ultimately, our attempts to characterize choice with a generalized expected utility framework did not yield a model with rates of descriptive accuracy substantially exceeding 50 percent. Moreover, most models imply a high degree of conservative choice. We explore additional explanations, and attempt to rule out potential confounds, for the observed conservatism through an online goal choice paradigm.

Table 2.7 · GoalQuest-style menus offered in lab paradigm

	Grids solved	Award (¢)		Grids solved	Award (¢)
Menu 1			*Menu 2*		
Goal 1	6	10	Goal 1	4	5
Goal 2	8	20	Goal 2	6	20
Goal 3	10	35	Goal 3	12	35
Menu 3			*Menu 4*		
Goal 1	8	10	Goal 1	10	10
Goal 2	10	20	Goal 2	14	20
Goal 3	12	35	Goal 3	18	25
Menu 5			*Menu 6*		
Goal 1	6	10	Goal 1	4	5
Goal 2	8	20	Goal 2	6	10
Goal 3	10	35	Goal 3	8	20
Goal 4	16	40	Goal 4	10	30

Menus presented to participants in the lab paradigm. Goal thresholds are denomenated in the number of addition search tasks that can be completed in 4 minutes (see Appendix Figure A5). Each participant made a selection for each menu with the understanding that one menu would be selected at random to determine award payments.

2.6 Goal choice in experimental paradigm

To further investigate the motives for conservative goal choice, we designed and administered an experimental goal-reward paradigm in the context of an online effort task. The incentive-compatible paradigm was intended to resemble GQ but with lower stakes, a shorter evaluation period, and multiple decisions per subject. We supplemented the effort task with a series of elicitations that captured demographics, beliefs, and other decision-relevant factors including risk preferences and loss aversion parameters. The online paradigm permitted us to corroborate findings from the field with increased statistical power, assess additional decision-making theories, and rule out potential confounds.

2.6.1 Research design and protocol

We programmed the stylized goal-rewards program and effort task using the Qualtrics platform. Following a series of background questions, we informed participants that they would be participating in a timed effort task where they would be asked to solve as many grids as they could and compete for a financial reward. To solve a grid, subjects had to identify the unique pair of numbers within a 9-number grid whose sum equaled 10 (see Appendix Figure A5). After a brief opportunity for practice, participants were introduced to the goal-rewards program, called GoalQuest, via a webflow resembling the actual webflow used by real-life employees.

The web-flow described the rules of the goal-reward program, the structure of the four-minute effort task, and the all-or-nothing goal menu. To increase the statistical power of our choice characterization, instead of then asking participants to select a performance goal from a single menu, we asked them to select a single goal from each of six distinct menus, explaining that one menu would be randomly selected to determine their potential bonus. We additionally elicited participant forecasts about their performance and tested their comprehension of the goal-reward program. Regarding the former, we elicited beliefs on the likelihood of solving 4, 6, 8, 10, 12, 14, 16, and 18 grids during the four-minute performance period to impute an expected overall performance for each participant, a slight departure from the field where we directly elicited beliefs about attaining each goal (see Appendix A5 for more on how imputations were made). Finally, at varying points, we assessed several decision

Table 2.8 · Goal choice, beliefs, and goal attainment: online paradigm

		Sample restricted by goal choice		
	All	**Goal 1**	**Goal 2**	**Goal 3**
Panel A. Goal Choice				
Goal 1	0.43			
Goal 2	0.32			
Goal 3	0.25			
Number of subjects	277	207	201	123
Number of choices	1108	471	356	281
Panel B. Employee beliefs				
Expected performance	11.0	8.5	11.3	14.8
Expected/actual performance	1.5	1.4	1.4	1.6
Panel C. Goal Attainment				
Goal 1	0.68	0.52	0.78	0.84
Goal 2	0.51	0.28	0.60	0.76
Goal 3	0.27	0.08	0.23	0.64
Earned reward (average)	0.12	0.05	0.12	0.22
Earned reward (average) given goal attainment	0.20	0.09	0.20	0.35

This table summarizes goal choice, beliefs, and goal attainment overall and separately by goal choice for participants in the lab paradigm. Only data from choices on three-goal menus are included; choices on four-goal menus are excluded. Each datapoint comprises a single choice made by a participant, including on menus that were later not chosen to be incentivized. The experiments asked each participant to indicate their goal-choice across six distinct menus in the context of an online effort task. Specifically, Panel A summarizes goal choice, Panel B summarizes actual and expected performance, and Panel C summarizes goal attainment and earned rewards. The data on goal attainment and rewards are hypothetical because most choices made by participants ended up not being incentivized. To calculate the data in Panel C, we treat all of the six menus as incentivized for each participant.

making parameters including risk aversion (using hypothetical lotteries adapted from Eckel, El-Gamal and Wilson 2009), a taste for competition, decision anxiety, and beliefs about relative grid-solving ability.

We strategically designed the six menus to facilitate tests of mechanisms (see Table 2.7). Informed by a series of pre-tests to gauge performance on the effort task, we designed a baseline menu (Menu 1) with goal levels (6, 8, 10) such that attainment would roughly resemble goal attainment in the field. The menu also featured non-linearly increasing rewards ($0.10, $0.20, $0.35). Three additional menus varied the relative attractiveness of Goal 3 (Menus 2 and 4) and the overall difficulty of the goal menu. The final two menus, which were designed to increase the robustness of the benchmark assessments and to test two additional heuristic choice theories, appended the options from the baseline menu with a relatively unattractive high goal option (Menu 5) or a relatively unattractive low goal option (Menu 6).

2.6.2 Comparison of goal choice in the lab and the field

We administered the online paradigm in May 2019 to 407 U.S. adults recruited from Amazon Mechanical Turk. We were compelled to discard data from 130 participants who had incomplete or internally inconsistent beliefs, leaving us with a final sample of 277 participants who made 1,662 goal choices. Few participants failed the comprehension check, and almost all that did passed on their second attempt. Table 2.8 summarizes the goal choices, beliefs, and performance of participants for the four menus with a standard, three-goal, structure. While

Table 2.9 · Goal choice characterization for non-standard benchmarks: online paradigm

	Risk-neutral expected utility RE, $r=0$	**Subjective expected utility** $r=0.0003$	**Fuzzy subjective expected utility** +/- 5% error	**Rank-dependent expected utility** Prelec $\alpha=\beta=0.65$	**Consumption + gain-loss utility** RP = g+1; η = 3	**Alternative heuristics** **Contextual sorting**	**Taste for competition**
Panel A. Optimal choice							
All Menus (6/6)	0.13	0.04	0.15	0.03	0.18	0.10	0.09
Nearly All Menus (5+/6)	0.29	0.16	0.43	0.13	0.40	0.12	0.10
All 3 Goal Menus (4/4)	0.15	0.16	1.00	0.09	0.25	0.13	0.12
All 4 Goal Menus (2/2)	0.35	0.10	0.15	0.10	0.40	0.15	0.15
Panel B. Conservative choice							
All Menus (6/6)	0.08	0.08	–	0.17	0.03	0.14	0.15
Nearly All Menus (5+/6)	0.23	0.25	–	0.39	0.10	0.21	0.23
All 3 Goal Menus (4/4)	0.09	0.09	–	0.17	0.04	0.17	0.25
All 4 Goal Menus (2/2)	0.41	0.62	–	0.62	0.17	0.18	0.16

This table characterizes the share of optimal (Panel A) and conservative (Panel B) goal choice for experimental participants under a range of standard and non-standard benchmark models. The experiments asked each participant to indicate their goal-choice across six distinct menus in the context of an online effort task. The first column characterizes choice for a baseline model of risk-neutral expected utility assuming rational expectations. The remaining columns all assume subjective beliefs and characterize choice for, from left to right, a subjective expected utility model assuming moderate risk aversion ($r=0.0003$), a fuzzy expected utility model that permits 10 percent +/- errors in calculations of expected utility ($r=0.0003$), a rank-dependent model of expected utility that applies nonlinear weights to potential outcomes using the Prelec (1998) function, and a model of composite gain-loss utility with the goal + 1 reward as the reference point and η = 3. A final set of columns characterizes choice for a set of heuristic-choice models whose details we describe in the text. The blank cells in Panel B reflect the inability to uniquely characterize conservative choices for the benchmark involving fuzzy expected utility.

participants selected Goal 3 at a slightly lower rate than in the field, they exhibited substantial overconfidence about their performance, and similar levels of attainment for Goal 3 (notably, attainment was somewhat higher for the lower goals). Critically, the experimental data corroborates the pattern of conservative choice relative to the expected value benchmark. Overall, participants chose Goal 3 only 25 percent of the time, despite it maximizing expected value in 51 percent of instances (based on either subjective beliefs or rational expectations, calculated with the same procedure used in the field). Restricting focus to the menus resembling GQ most closely (Menus 1 and 3), participants chose the high goal only 33 percent of the time despite it still maximizing expected value in 69 percent of instances.

2.6.3 Characterization of choice

We proceed to characterize choice by first assessing selected benchmark models from the field analysis. Because we asked each participants to make six distinct choices, we characterize choice by reporting the share of participants whose choices across multiple menus adhered to the predictions of the given benchmark. This strategy not only permits greater statistical clarity as to a given participant's behavior, it also allows us to incorporate some error into the characterization – that is, we can assess whether a large share of participant decisions nearly, but not completely, adhere to some benchmark model. Accordingly, Table 2.9 characterizes choice for the risk-neutral EU benchmark as well as benchmarks incorporating subjective beliefs, moderate risk aversion, non-linear decision weights (using the Prelec 1998 function) and gain-loss utility (using one of the most successful formulations from the field analysis). The table suggests that none of these models adequately explain participant choice. The most successful of the four benchmarks fully explains the choices of only 18 percent of participants. Even allowing for a small margin of error, in the form of 5 of 6 menus, the benchmark models can explain at most 40 percent of participant decisions. To consider whether the descriptive inaccuracy of the models may be due to the presence of computational imprecision in the evaluation of choices from a specific menu, we further assess a "fuzzy" EU model that rationalizes any decision that is within an arbitrarily set +/- 5 percent error interval. Such a model can explain 43 percent of participants' decisions, when explanation is measured by 5+ correct choice predictions out of 6 menus.[3]

Finally, we consider two heuristic models of choice that have been asserted in the literature as possible explanations for financial choices from a menu of options and that we could not assess from the field data. The first of these heuristics is that of contextual sorting, the presumption that individuals, otherwise uncertain of what goal to choose, select a goal that corresponds to their perceived standing in a relevant distribution, in this case productivity. Contextual sorting was suggested as a potential explanation for 401(k) allocation choices among employees (Kamenica 2008). The heuristic could be rationalized only if the menu was constructed in such a way that low goals maximized the expected utility of low-performing participants and high goals maximized the expected utility of high-performing participants. We tested for this heuristic by asking participants to assess their grid-solving ability relative to other participants and then assessing whether this relative ranking maps to choices from the menu (i.e., high ability participants select goal 3 and average ability participants select goal 2).[4] The second heuristic we test is predicated on taste for competition. Presuming that individuals might sort into a goal based on how relatively competitive they perceived themselves, we asked participants to rate their relative competitiveness and then mapped responses to goal choices on each menu.[5] The table indicates that neither of the two heuristic models substantially helped to explain goal choice.

[3]To implement the fuzzy model, we consider a choice to be a successful prediction by the model if it is within 5 percent of the utility of the model's predicted choice.

[4]Specifically, we asked: "relative to the other people in this study, how do you think the number of grids you solve will compare?". Participants responded on a five-point scale from "well below average" to "well above average." Our simple model predicts anyone who chooses "well below average" or "below average" will select Goal 1, anyone who chooses "average" will choose Goal 2, and the remainder will choose Goal 3. (On four-goal menus, we predict only those who report "well above average" will choose Goal 4.)

[5]We elicited relative competitiveness on a five-point scale ranging from "much less competitive" than other people to "much more competitive" than other people.

2.7 Conclusion

We describe new evidence on the magnitude of financial risk-taking and its underlying motives. Our evidence describes the decisions of several thousand employees in the context of a popular employee reward program. We saw this setting as uniquely helpful for understanding risky choice given the diversity of the decision makers, the wide-ranging financial magnitudes, the simplicity of the choice environment, and our visibility into contemporaneous employee beliefs. Our initial finding is to document substantial risk aversion in the goal choices of employees. From the perspective of an expected utility benchmark, with rational expectations, these decisions resulted in an average expected foregone reward of $139 and a maximum expected foregone reward of $1435. The excess conservativism of employees, as judged from the rational expectation or perfect information benchmark, persisted in decisions with large financial stakes and among highly tenured employees (10+ years of experience). We proceed to show that conservative goal choice cannot be explained by the diminishing marginal utility of employee wealth or by common behavioral departures from the expected utility framework such as biased beliefs, non-linear decision weights, or gain-loss utility (using an expansive set of potential reference points and functional specifications). Across the models we tested, none explained substantially more than one-half of employee choices.

To generate additional evidence on mechanisms, and to rule out potential confounds, we designed and administered an experimental incentive-compatible rewards program, resembling GQ, in the context of an online effort task. The experimental paradigm permitted us to directly elicit person-specific decision-making parameters (e.g., beliefs, risk preferences, loss aversion, taste for competition, emergency liquidity), confirm understanding of program rules, and observe multiple goal choices per subject across six strategically varying menus. While stylized and involving far smaller stakes than in the field, the online paradigm produced the same general pattern of conservative goal choice and substantial overconfidence about productivity as we had observed in the field. And as with the field data, we found that one cannot adequately explain the goal choices of experimental subjects using the expected utility framework, even after incorporating plausible levels of risk aversion, subjective beliefs, non-linear decision weights, or gain-loss utility. We additionally assessed, and then rejected, two alternative mechanisms that have been suggested in the literature as potential explanations for menu-based choice involving contextual sorting and a taste for competition.

Ultimately, the evidence across the field and lab implies that the standard expected utility framework, even after modification to include the departures most seriously contemplated by the literature, struggles to explain the simple, yet consequential, choices of employees in this setting. Either choice reflects substantial heterogeneity across the motives we considered or it reflects some motive that we did not consider such as some heuristic choice strategy that incorporates the decision anxiety that financially meaningful decisions might trigger. One practical implication of this research is that if conservative decisions in other domains, such as insurance or asset markets, do not reflect risky-preferences, then welfare estimates of policies in those domains may not be accurate. Moreover, there may be scope for improving the optimal design of programs and policies that compel individuals to engage financial risk We hope that further work can clarify the specific decision process that underlies choice in this setting and to confirm whether the same decision processes might help to explain apparent risk aversion in other consequential settings.

Chapter 3

Do risk perceptions change after a disaster? Evidence from flood-prone counties

3.1 Introduction

The tendency for insurance demand to spike after disasters has been observed broadly across different risk domains and countries (Palm 1995; Browne and Hoyt 2000; Gallagher 2014; Kousky 2017). New disasters presumably convey some information about future disaster risk, and swings in demand could in principle be attributed to consumers updating their beliefs about future risk within a rational-expectations framework. Under this interpretation, consumers are carrying an optimal level of insurance both before and after flooding events.

However, the size of the demand spikes observed in the wake of flooding and hurricanes, for example, are hard to reconcile with the prevailing scientific view that one extreme flooding event in a given location is not predictive of similar events in the same location in the near future (Gallagher 2014). Additionally, evidence from Gallagher (2014) and Kousky (2017) indicates that insurance takeup falls back to baseline levels quite quickly in the years after a disaster. These patterns are hard to reconcile with a rational expectations model where consumers make small adjustments to risk perceptions using Bayesian reasoning.

The literature has offered at least two competing explanations. First, lab evidence shows that people can switch between mental states of not worrying at all and worrying "too much" when it comes to insurance purchases (McClelland, Schulze and Coursey 1993). Under this hypothesis, coastal residents are underinsured most of the time, but dramatically overinsured in the years after an event. The second alternative explanation is that consumers are usually inattentive to unusual risks like floods, but are galvanized by major flooding events to consider their risks anew and make a considered insurance decision. This would suggest that flood-prone households are usually underinsured, and only become adequately insured after a disaster focuses their attention on the threat.

Both explanations indicate that policy interventions that help people develop beliefs consistent with actual risk may be welfare-improving. However, less is known about changes in beliefs, both in the immediate aftermath of a disaster and as the disaster recedes. In this paper, we use longitudinal survey data to understand how beliefs about risk evolve after two different treatments: exposure to FEMA flood maps and exposure to a local severe flooding event.

To measure the responses, and in the former case to administer the treatment, we conduct a survey of 1,582 coastal U.S. residents using Amazon's Mechanical Turk platform. We elicit information on respondents' beliefs about the likelihood of home flooding and its potential cost, their level of "worry" about home flooding, and their risk preferences. We also include questions about willingness to pay for hypothetical flood insurance policies and actual flood insurance takeup. The first treatment is an experiment in the survey wherein some participants are randomly selected to view online FEMA flood risk maps covering their hometowns and are given brief instructions in how to read them. This experimental treatment allows us to estimate the effect of neutral information disclosure on flood beliefs and willingness to pay for insurance.

The second treatment is generated by our longitudinal study of the same respondents. By comparing responses before and after hurricane season, and across areas that are differentially impacted by hurricanes during the 2016, 2017, and 2018 seasons, we are able to estimate the effect of flood exposure on our three variables of interest (perceived risk, worry, and willingness to pay) using a difference-in-difference procedure.

This paper contributes to the literature on insurance demand and learning about risk in several ways. Past studies of insurance demand response to floods, notably Gallagher (2014) and Kousky (2017), are based on aggregate data and do not measure individual decisions or underlying beliefs. First, our cross sectional analysis provides descriptive evidence of what factors are associated with increased perceived risk, worry, and insurance willingness to pay. One interesting finding is that being a climate change believer is associated with both higher worry and higher perceived risk. This finding underscores the importance of educating the public about climate change and its relationship with flood risk.

Second, the experimental treatment, exposure to FEMA flood maps, enables us to measure the causal impact of a more neutral shock and compare the effect to flood exposure itself. We find that map exposure significantly reduces perceived flood risk on average, and significantly strengthens the relationship between perceived risk and insurance demand, but has differential impact on respondents living at different elevations (and hence with different levels of risk). Further, the lower perceived risk on average is not consistent with actual risk. The results from this experiment demonstrate the potential for gains from better communication of flood risk.

Finally, with longitudinal survey data from approximately half of our sample ($n = 725$), We are able document the causal impact of flood exposure on beliefs about flooding likelihood and costs. We find that a flood shock does lead to a slight increase in perceived risk and worry. Moreover, we find that perceived risk and worry attenuate quickly as the flood experience becomes more distant. This result is consistent with both of the alternative explanations: both that people became more attuned to risk and their beliefs about a known risk changed.

The remainder of this paper is organized as follows. Section 2 provides a literature review and background on the question of insurance demand response to disasters. Section 3 describes the survey design in detail and presents descriptive statistics about the sample of responses collected. Section 4 details the empirical approach and Section 5 presents the results in graphical and regression form. Section 6 concludes with a discussion of the results and policy implications.

3.2 Background

Empirical research has found that rare events like earthquakes, flooding and cancer clusters can induce behavioral changes that are suggestive of dramatic shifts in underlying beliefs. Several studies have analyzed behavior after such events, ranging from insurance purchase decisions (Palm 1995; Gallagher 2014; Said, Afzal and Turner 2015) to real estate transactions (McCoy and Walsh 2018; McCoy and Zhao 2018; Bin and Landry 2013; Davis 2004) to resource allocation by corporations and governments (Dessaint and Matray 2017; Wibbenmeyer, Anderson and Plantinga 2016). The general tendency is for people to behave in ways consistent with either increased perceived risk, more risk-averse preferences, or both, such as buying earthquake or flood insurance after an earthquake or flood occurs nearby. Most of these papers do not have disaggregated microdata on individual decisions, except for Dessaint and Matray (2017) and the real estate transaction studies.

The clearest evidence for the insurance demand response to flooding events comes from Gallagher (2014), who assembles a panel dataset of U.S. flooding events and insurance takeup running from 1980 to 2007 and conducts an event study. He finds that a flood disaster declaration in a given county increases participation in the National Flood Insurance Program (NFIP) in that county by an average of 9%, but that the effect diminishes and eventually disappears over the ensuing decade. In a similar study, Kousky (2017) finds a slightly smaller effect on flood insurance takeup due to hurricanes, and finds that it fades out even more quickly.

Flood insurance is an ideal context to study demand responses because flood insurance prices are set by the government-run National Flood Insurance Program (NFIP) and do not generally change systematically in response to flood events (Gallagher 2014). However, in both studies the researchers rely on aggregate insurance data, so they can not analyze mechanisms on an individual level nor do they use any data on perceptions of flood likelihood or costs, or individual risk preferences.

Another strand of the literature has investigated changes in risk preferences after catastrophic events. Several papers measure risk preferences among populations recently exposed to disasters like Hurricane Katrina or the 2004 Indian Ocean tsunami. These have found varied results (Cassar, Healy and von Kessler 2017; Cameron and Shah 2015; Bchir and Willinger 2013; Eckel, El-Gamal and Wilson 2009). Only Hanaoka, Shigeoka and Watanabe (2018) include longitudinal data from before and after a major disaster, and they find increased tastes for risk among those more exposed to the 2011 Fukushima earthquake.

Very few studies have used explicit data on beliefs about disasters. Bakkensen and Barrage (2017) conduct a door-to-door survey of coastal residents in Rhode Island and find a negative correlation between perceived flood risk and actual risk as estimated by U.S. government authorities, providing partial inspiration for the present work. A study by Deryugina (2013) analyzes how localized temperature fluctuations affect beliefs about climate change and finds that the beliefs of people with conservative political attitudes are responsive to sustained temperature fluctuations. Royal and Walls (2019) is perhaps the closest parallel to our approach: they study the relationship between risk perceptions and insurance take up and find that flood plain residents are over optimistic about their risk. Other researchers have looked at data on beliefs about rare events in a non-learning context (Reynaud, Aubert and Nguyen 2013; Cameron 2005; Champ, Donovan and Barth 2010).

Finally, a literature on risk aversion in the presence of ambiguity finds that risks with an ill-specified probability of occurring (like having precisely 13 rainy days in a given city next month) are sometimes treated differently than ones that are numerically specified, as they are, for example, in the context of government flood maps (Di Mauro and Maffioletti 2001; Schade, Kunreuther and Koellinger 2012). Additionally, Schade, Kunreuther and Koellinger (2012) find that self-reported "worry" is more important than subjective probability in predicting insurance demand, motivating us to include a worry elicitation question in our survey.

3.3 Survey and data sources

3.3.1 Survey construction

We conducted the initial survey on Amazon's Mechanical Turk platform, collecting responses between June 4 and July 31, 2018. Responses were solicited from 124 coastal counties in 13 states spanning the Gulf coast and most of the Eastern Seaboard. We had two goals in mind when determining the set of counties included in the study. First, we wanted to focus on people living near flood zones. These individuals are making the most high-stakes decisions about flood insurance coverage and are presumably most sensitive and attuned to the threat of flooding. Second, we want to focus on regions with particular vulnerability to hurricanes, as quasi-experimental exposure to hurricanes is the basis of our most important identification strategy. While narrowly targeting the survey to coastal residents potentially weakens the external validity of our findings, we argue that the coastal sample is drawn from the true population of interest, and that any differential response to the two treatments we consider by people living far from flood zones is not of first-order significance.

Records from the past 150 years suggest that coastal areas on the Gulf of Mexico, the Atlantic coast below New York City, and Long Island are the areas of the mainland United States most frequently targeted by powerful hurricanes (Blake, Landsea and Gibney 2011). The counties included in the study region were selected based on their proximity to the shore and, in a few cases, unusually high historical flood insurance claims. All counties along the Atlantic coast from Texas to Long Island were included in the region. Several counties in Texas, Louisiana, and Florida that were not coastal but which had historical claim rates since 1978 greater than 0.02 per capita were also included (Casey 2017). Additional counties and city jurisdictions in Maryland, Virginia, and Delaware

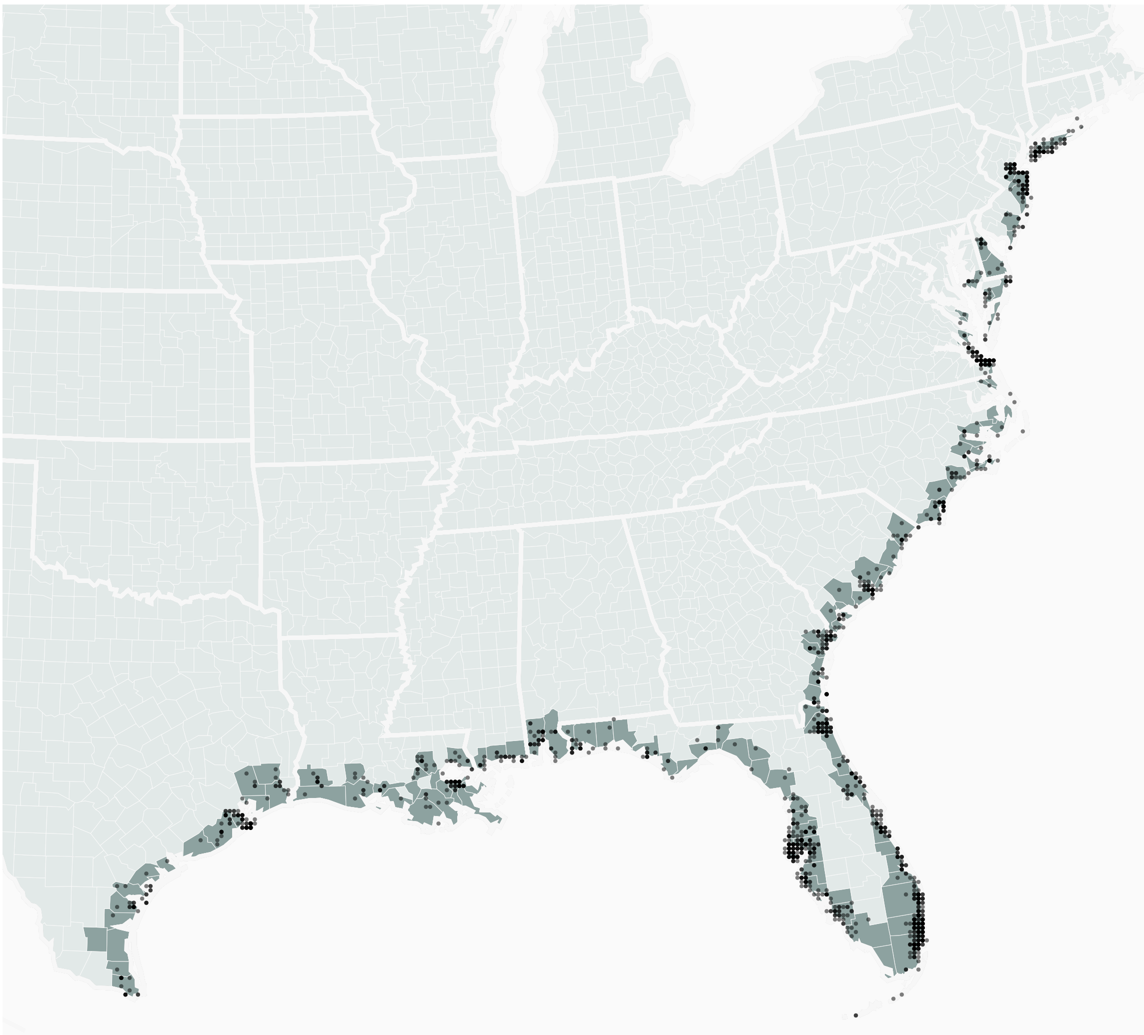


Figure 3.1 · Location of 1,582 survey responses

The figure above shows the approximate geographic location of each of the 1,582 survey respondents within the United States. Darker points indicate a greater density of responses. The 124-county coastal region where the survey was fielded in shaded dark gray.

which lie on the Chesapeake Bay or Delaware Bay were included if they met this threshold as well. Harris County (Houston) qualified for inclusion but was specially excluded due to the potential for its large population to distort the sample, but many coastal counties in Texas with exposure to Hurricane Harvey in 2017 were included.

Figure 3.1 shows the 124-county region (highlighted) and the geographic distribution of the 1,582 valid responses received during the initial sample window. As expected, responses are clustered in the most densely-populated parts of the region, including Galveston, New Orleans, Tampa Bay, Miami, Norfolk/Virginia Beach, and the New York City metropolitan area. Approximately 44% of the sample is concentrated in the state of Florida.

The survey included questions on basic demographic characteristics, a lottery-choice elicitation of risk preference (Charness, Gneezy and Imas 2013), as well as questions about respondents' personal history with flooding and the situation of their home relative to the water. See Appendix A6 for a complete list of survey questions.

In 2019, we sent follow-up survey initiations to all participants. During our second round of surveys, we received 725 valid individual responses, giving us a 46% response rate for our longitudinal study. We find that the responders and nonresponders were closely balanced on observable characteristics, reducing concerns about response bias within our initial sample.

3.3.2 Additional data sources

In addition to personal survey data on beliefs and preferences, we collect information about local flood risk that we can then match to respondents based on their reported municipalities. The richest data on local flood risk comes from the dataset analyzed by Gallagher (2014), which has information at the municipality level. These include the number of flooding events between 1990 and 2007 and average NFIP monthly premium paid in 2007.

Local flood risk data will allow us to analyze the relationship between subjective perceived flood risk and objective historical data that should be closely linked to flood risk, and how that relationship may change under various treatments. However, we can only match 82% of our sample to the local flood risk data, as certain communities are too small to be broken out separately in the data and some larger communities are excluded for idiosyncratic reasons (e.g. New Orleans, Louisiana). To complement this data, we also gather county-level disaster declaration data from the Federal Emergency Management Agency that extends from 1965 to 2016 that covers virtually every county in the United States and the entirety of our sample (Federal Emergency Management Agency 2016).

We also include information on local participation in, and activities performed in service of, the Community Rating System (CRS) administered by the National Flood Insurance Program. Communities that participate in CRS can earn points based on activities performed that abate flood risk; residents in high-scoring communities can receive discounts on flood insurance up to 45% (Chapter 3, Federal Emergency Management Agency 2017). Data provided by FEMA lists the counties and municipalities participating in CRS as of October 1, 2017, and their point totals for various activities. This allows us to determine which localities are receiving credit for providing a "Map Information Service" to residents.

Additionally, we use the Google Earth API to impute home elevation based on the elevation at the geocoded location for each ZIP code.

3.3.3 Map exposure experimental design

To test the effect of exposure to authoritative flood risk information, we randomly assign a subsample of the survey respondents to complete an additional section of the survey. Respondents in this condition were directed to find their homes on online flood hazard maps maintained by FEMA at `https://msc.fema.gov/portal/home` before they answered the key questions about perceived flood risk, flood worry, flood cost, and willingness to pay for insurance. The data underlying these maps is used to determine premiums for flood insurance available through the NFIP (Gallagher 2014).

The maps available online feature two brightly-colored zones: an orange zone indicating the "500-year flood zone" corresponding to a $\geq$0.2% annual chance of inundation and a blue zone indicating the "100-year flood zone” corresponding to $\geq$1% annual risk (see Figure A1 for an example of how these maps appear in a web browser). Respondents were given a brief primer on the flood zone definitions and were asked to look at the maps to see if their home or any homes nearby were in either the 500-year or 100-year flood zone. They were also asked to indicate qualitatively how their beliefs about flood risk changed upon seeing the map.

We tailored the treatment as narrowly as possible to serve as a simple exposure to flood maps without any extraneous impacts. But there is always the possibility that the acts of searching for one’s flood map online and reading our brief explanation of the flood zone definitions could have had their own effects.

We perform several placebo tests by assessing the “effect” of map exposure on past events that could not possibly be affected by experimental assignment at the time of the survey, such as whether the respondent has already purchased flood insurance or has had a past flood in her current home. Additionally, we test the effect on attitudinal questions that we hypothesize should not have been affected by exposure to flood maps. Finally, we test whether respondents anchored on the two numerical percentages mentioned in the brief explanation of flood zones: 1% and 0.2% (see Appendix A6, Question 20). Especially since these numbers were referring to one-year risk levels and the question asked respondents to estimate ten-year flood likelihoods, the inclusion of these numerical values might have confused respondents and induced them to think about one-year risk levels. The results are available in Appendix Table A2 and show that the map exposure treatment generally did not have unexpected consequences, with the exception of a marginally significant effect on ability to name major past floods or hurricanes that struck one’s city or town.

3.3.4 Sample composition

Table 3.1 displays descriptive statistics for the resulting sample. Due to the voluntary nature of the online survey approach, the sample is not intended to be representative of the population in the covered counties, let alone the U.S. population at large. As expected, the sample skews quite young, with approximately 58% of the sample under age 35. These people are approaching their prime home-buying years, so their perspective on flood insurance may be of particular interest. The sample is better-educated and quite likely to agree with the notion that climate change is occurring, but relatively low-earning given its high educational attainment (possibly owing to age or current enrollment in school).

In an effort to keep average response time for the survey low, we assigned only 25% of the sample to the longer treatment condition. This yielded a substantial experimental group sample of 365 valid responses. While precisely one quarter of the sample was assigned to the treatment condition, only 23.07% of the final sample is treated. The discrepancy is accounted for by slight differences in attrition rates, presumably because respondents in the experimental group were asked to complete a slightly longer survey. Table 3.1 shows that the experimental arm and the control group are quite well-balanced on the observable characteristics. Nonetheless, there could be some small latent differences introduced by differential attrition.

By design, the sample is much more exposed to the threat of flooding than the American population on average. Many respondents report living within twenty minutes’ walking distance of a body of water, like the ocean or a river, that could flood their homes. Many are living at very low elevations, with 44% of the sample in homes at an elevation of 10 feet above sea level or less. Most can remember or name past hurricanes that struck their towns of residence, 49% report carrying flood insurance policies, and nearly 8% have experienced flooding in their homes in the past year.

Table 3.2 displays distributional information on the four key response variables elicited in the survey. These data reinforce the sense that this sample is particularly flood-prone, and regards itself that way. The mean perceived 10-year flood risk is 3.34%, and roughly a quarter of respondents perceive 10-year flood risk at greater than 10%, consistent with a home in the 100-year flood plain. By comparison, Crowell et al. (2010) estimate that only 3% of the U.S. population lives in 100-year flood plains as designated by the FEMA maps.

Table 3.1 · Sample composition

Variable	Full sample *(n = 1,582)*	Experimental group *(n = 365)*
Survey data		
Female	0.58	0.57
White/Caucasian	0.74	0.76
Age		
under 35	0.58	0.58
over 55	0.10	0.10
College degree	0.51	0.48
Income greater than $60k	0.30	0.29
Homeowner	0.54	0.52
Lives in single-family, detached home	0.62	0.63
Identifies as Democrat	0.37	0.39
Believes climate change is happening	0.80	0.79
Can remember major hurricane	0.80	0.76
Risk preferences		
chose least risky lottery	0.45	0.47
chose most risky lottery	0.09	0.08
Past flood in current home	0.17	0.15
Home flooded in past year	0.08	0.07
Walking distance to water	0.53	0.50
Carries flood insurance	0.49	0.46
Map information service (CRS)	0.69	0.66
Years in current home	6.79	7.06
	(7.56)	(8.04)
County flood events since 1965	4.38	4.25
	(2.67)	(2.62)
Home elevation (ft)	24.25	23.43
	(34.91)	(30.33)
Worry about asteroid (1-10 scale)	2.12	2.00
	(1.89)	(1.79)

Sample mean values with standard deviations in parentheses for continuous variables. Home elevation values imputed for some respondents based on municipality or ZIP code. See Appendix A6, Question 12 for the risk elicitation language.

Table 3.2 · Distribution of key response variables

Variable	Mean	Std. dev.	10th pctile	Median	90th pctile
Perceived 10-year flood risk	3.30%	5.11%	0.03%	0.94%	10.98%
Worry about home flooding (1-10 scale)	4.46	2.51	1	4	8
Perceived cost of flood in home	$61,860	$250,974	$3,111	$25,000	$120,000
WTP for flood insurance	$816	$1,183	$120	$360	$1,800

Full-sample means, standard deviations, and percentiles for response variables. Perceived costs were determined by asking respondents to estimate the damages from a one-foot flood in their homes. Willingness to pay for insurance was elicited with a multiple price list with 16 options. Willingness to pay was expressed in terms of an annual premium for a policy that covered $100,000 in damages.

Table 3.3 · Problematic observations in survey sample

Issue	Count	Share
Total responses collected	1,716	
Did not appear to live in eligible county	134	0.08
Admissible sample	1,582	0.92
Full usable sample	1,582	
Completed survey in under 5 minutes	83	0.05
Failed flood depth attention question	346	0.22
No disaster history for municipality	269	0.18
Municipality was imputed	38	0.02
Home elevation was imputed	385	0.24
Inverted multiple price list	46	0.03
Multiple preference reversals in MPL	153	0.10
Missing or unreliable demographic data	17	0.01
None of the problems listed above	749	0.47
Map experimental group	365	
failed attention question	114	0.31
badly failed attention question	43	0.12

Municipality imputed based on home ZIP code or inferred from reported municipality in cases of misspelling, etc. Home elevation imputed based on ZIP code for respondents who reported an implausible elevation (see Section 3.5.4 for a more detailed explanation). Respondents with "inverted" multiple price lists reported that they were not willing to pay small amounts for the insurance contract, but were willing to pay larger amounts, apparently reversing the two responses (see Appendix B, Question 29). Respondents with "multiple preference reversals" had internally inconsistent answers to Question 29 and their WTP data was discarded. See Appendix B for attention questions: Question 26 is the flood depth attention question and Question 22 is the attention question for the experimental group.

3.3.5 Drawbacks of the internet survey approach

While online survey platforms present an opportunity to reach a broad population living across a wide geographic area – a key factor in this study – they also have some important limitations (Paolacci, Chandler and Ipeirotis 2010). One major drawback of online survey-taking is that respondents are answering questions in private and not in a supervised laboratory environment free of distractions. Respondents may be working on multiple other tasks online and off, and have monetary incentive to move quickly through the questions but little incentive to provide truthful answers. In the context of this survey, for which accurate geographic information is crucial to evaluating flood risk and for assigning the treatment variable in the case of the hurricane exposure analysis, the inability to verify respondents' location is potentially comprising. While we discard respondents who listed non-existent ZIP codes or whose ZIP codes and municipalities were mutually inconsistent, we cannot be sure that respondents are actually living in the ZIP codes and municipalities that they claim.

We include two attention questions to test whether respondents are rushing through the survey without carefully considering the questions or their responses. Respondents must be paying attention to the survey and the attention question specifically to record the right answer. Overall, the performance on these questions was somewhat poor, with 21% of the sample failing to answer a question about the depth of floodwaters asked about in a previous question, and 31% of the map experimental arm sample failing to answer a question about which color on the flood map zone corresponded to the highest flood risk. Table 3.3 details these and various other potential problems with some observations in the sample. In Section 3.5.4, we perform robustness checks where we rerun the main analyses presented below with problematic observations excluded.

3.4 Approach and hypothesis

3.4.1 Cross sectional analysis

We start by analyzing determinants of our three variables of interest: risk perceptions, levels of worry, and willingness to pay for insurance. These analyses do not have a causal interpretation, but do allow us to document descriptive correlations between characteristics of our survey respondents and the response variables.

For this analysis, we run the following regression:

$$Y_i = \alpha + \beta \mathbb{I}\{\text{map exposure}\}_i + \delta \mathbf{X}_i + \gamma \mathbf{Z}_i + \eta_j + \epsilon_i, \tag{1}$$

where

- $\mathbf{Y}_i$ is one of our three main outcomes
- $\mathbf{X}_i$ are individual-level hydrological controls like home elevation and local flood history,
- $\mathbf{Z}_i$ are demographic controls like age and education, and
- η_j is a fixed effect at the municipality level,
- ϵ_i is an error term,

We expect proximity to water will be positively associated with risk, worry, and willingness to pay for insurance. We also expect flood experience to be positively correlated with all our dependent variables, however we expect more recent floods to have a stronger effect. Whereas, we expect elevation to be negatively correlated with all three variables of interest. Finally, we hypothesize that being a self-reported climate-change believer is positively correlated with our variables of interest.

3.4.2 Map exposure treatment

The most basic question that our experiment can answer is whether exposure to FEMA flood maps causes people to revise their risk perceptions, worry levels, and willingness to pay for insurance. We take advantage of the crisp identification provided by our randomized controlled trial to estimate the following model of the impact on outcome Y for individual i in municipality j using OLS:

$$Y_i = \alpha + \beta \mathbb{I}\{\text{map exposure}\}_i + \delta \mathbf{X}_i + \gamma \mathbf{Z}_i + \eta_j + \epsilon_i, \tag{2}$$

where

- $\mathbf{X}_i$ are individual-level hydrological controls like home elevation and local flood history,
- $\mathbf{Z}_i$ are demographic controls like age and education, and
- η_j is a fixed effect at the municipality level,
- ϵ_i is an error term,

and β captures the effect of random assignment to map exposure. We will estimate this effect for the four response variables listed in Table 3.2 above:

1. Subjective probability of a home flooding event in the next ten years
2. Self-reported worry on a 1-10 scale about a future home flooding event
3. Perceptions of flood costs, as measured by the following question: "If your home flooded to a depth of one inch, how much do you think it would cost to repair the damage?"

4. Willingness to pay for flood insurance

We do not have strong priors about whether exposure to the flood maps will raise or lower flood expectations on average. There is a possibility that respondents will be primed by the topic of the study or the FEMA flood map website, and will begin to grow alarmed about the possibility of floods. On the other hand, obtaining information that may be perceived as authoritative could be reassuring to some respondents, and lull them into a sense of security. Generally people underestimate the risk of rare events like floods, so if the maps serve to shift people toward a more hydrologically-informed view of their personal risk, flood perceptions should increase on average (McClelland, Schulze and Coursey 1993).

Of more central interest is how map exposure affect the *relationship* between home characteristics and flood perceptions, because this could illuminate the learning process of coastal residents. We do not elicit the response variables from participants in the map exposure treatment before they are treated – only after – and thus cannot analyze the map exposure effect at the within-subject level. However, we can identify the informational treatment effect by comparing the control group and experimental group.

To do so, we take the specification in Equation 2 and add interaction terms for the hydrological variables:

$$Y_i = \alpha + \delta \mathbf{X}_i + \gamma \mathbf{Z}_i + \rho \mathbf{X}_i \times \mathbb{I}\{\text{map exposure}\}_i + \epsilon_i. \tag{3}$$

The coefficients on the interactions terms ρ will indicate whether people who have been exposed to the FEMA flood maps have a different approach to the problem of estimating their risk, one that for instance might rely more on home elevation and less on distance to the water. This could indicate that the map exposure treatment is having subtle but important heterogeneous effects on respondents with different levels of fundamental flooding risk.

We remove the municipal fixed effects from this model because of the concern that these fixed effects will mask the crucial variation between observations in home elevation and distance from the water that are important to identifying the interaction effect. Additionally, in an interaction between the municipality fixed effects and the map exposure indicator, many observations would be the lone observation in their cells.

3.4.3 Hurricane exposure treatment

After the 2018 hurricane season, we resurveyed the same respondents included in this analysis to build a longitudinal dataset. Hurricane Florence was the most damaging hurricane in terms of flooding to strike the U.S. in 2018.[1] Florence affected counties in South and North Carolina and provides a clean source of quasi-experimental variation. To measure the causal effect of hurricane exposure, we use a difference-in-difference formulation. To test the impact on outcome Y for individual i in municipality j in time $t \in \{\text{pre, post}\}$, we estimate the following OLS model:

$$Y_{it} = \alpha + \delta \mathbb{I}\{\text{post}\}_j + \gamma \mathbb{I}\{\text{hurricane exposure}\}_j + \beta \mathbb{I}\{\text{post} \times \text{exposure}\}_t + \rho \mathbf{X}_i + \psi \mathbf{Z}_i + \eta_i + \epsilon_i, \tag{4}$$

where β is the coefficient of interest and the outcome variable Y tested will be one of the following:

1. subjective probability of a home flooding event in the next ten years,
2. current subscription to a flood insurance contract (binary variable),
3. willingness to pay for flood insurance,
4. self-reported worry on a 1-10 scale about a future home flooding event,
5. perceptions of flood costs, and

[1] Hurricane Michael caused catastrophic damage in a sparsely-population part of Florida, but this was primarily due to wind damage.

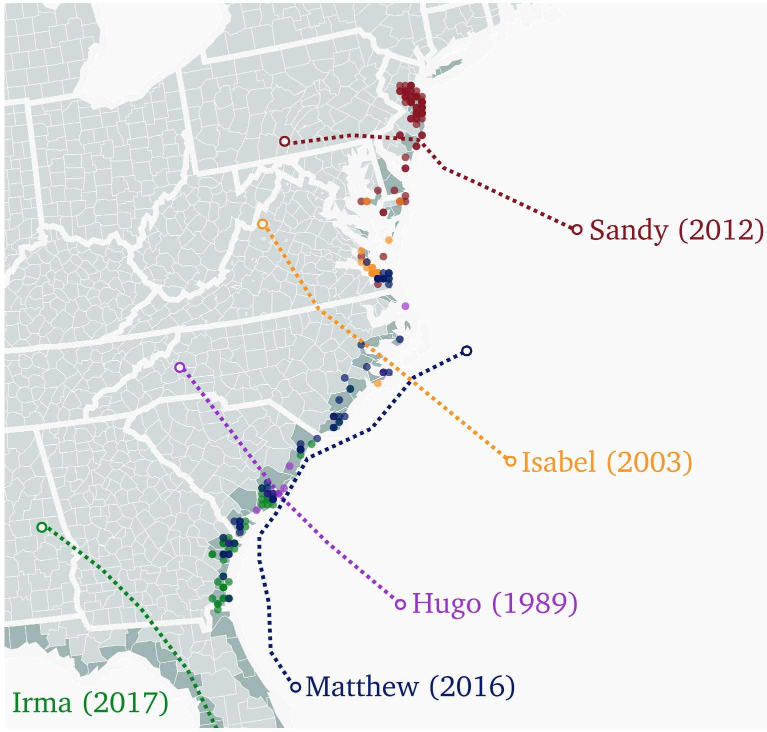


Figure 3.2 · Hurricane recollections

This figure displays the tracks of five major hurricanes from the past 30 years, along with the locations of respondents in seven states on the Eastern Seaboard who named one of these storms in response a question about "the last time [your city or town] experienced a major hurricane or major flood." Respondents who listed other storms or could not remember any major hurricanes are not shown (see Appendix A6, Question 35). Responses from New York and Florida were nearly unanimous and are not shown. Hurricane track data obtained from the NOAA Historical Hurricane Tracks tool provided by the National Oceanic and Atmospheric Administration.

6. risk preferences (6 risk-level bins).

We define the treatment as 1 for counties within the Presidential Disaster Declaration, and 0 otherwise, in the spirit of Gallagher (2014).

To study how perceived risks, beliefs, and insurance take up may attenuate the longer the storm recedes in memory, we also compare the changes in response across waves from those who experienced flooding in the years prior to our initial survey to all other respondents. In this case, the treatment is exposure to a hurricane in 2016 or 2017 (with Hurricane Harvey in Texas being the most prominent example).

Figure 3.2 displays some survey results that illustrate our identification strategy. It shows the tracks of recent major hurricanes that have impacted the Eastern Seaboard in the last 30 years, and the locations of respondents in seven states who reported that one of those five hurricanes was the last major hurricane to strike their hometowns as of summer 2018. Respondent points are color-coded to match the corresponding hurricane tracks. A few respondents in these seven states who reported a different hurricane or could not remember any major hurricane impacts are not shown.

The results are encouraging. Several respondents in South Carolina consider Hugo (1989) to be the last major hurricane to strike, while just a few miles to the south in Georgia most respondents consider themselves to have been struck by Irma (2017). The divergence in responses between South Carolinians and Georgians who live close to each other and presumably have similar long-term flood risk provides a source of variation in what could be a much more powerful learning experience than the brief experimental treatment in our survey.

3.4.4 Threats to causal inference

Identification of the two treatment effects we examine in this study relies on experimental and quasi-experimental identification variation, respectively. The map exposure treatment is applied to every fourth participant in the online survey, and the precise order in which respondents take the survey is presumably orthogonal to any other relevant factors, so the effects measured can reasonably be interpreted as causal. However, only 23% of the final sample – after incomplete and faulty responses are discarded – is from the experimental group, not 25% as would be expected if attrition rates were identical for both groups. The difference is small enough to be attributable to random chance, but could threaten to slightly bias the sample if experimental group participants with less perseverance were selected out of the sample.

Naturally, there are more threats to the validity of the quasi-experimental strategy for measuring the effect of flood exposure. First, defining a meaningful boundary for the treatment group may be difficult – people outside the official disaster area may still be impacted by a storm in a way that affects insurance demand, or it may be the case that only people near the center of the disaster zone were truly affected by a storm in the way we hypothesize. Second, the unavoidable geographical clumping of the treatment group means that there is likely to be a correlation between flood exposure in 2018 and flood exposure (or lack thereof) in recent years. For instance, if a new hurricane hits approximately the same region as Harvey, the effect will be different from a direct hit to an area like the Tampa Bay region that has not experienced a direct hit in living memory (Simmons and Neuhaus 2017).

Additionally, news coverage of the hurricane may blunt the gradient between people who are “directly” exposed based on where they live, and people who see video of floodwaters on cable news or on social media. Gallagher (2014) detects a flood insurance demand response in non-flooded counties that happen to share a local media market, and that increased coverage on local TV news, as measured by analysis of closed caption transcripts, was associated with increased response. In the case of a Katrina- or Harvey-level disaster that garners national news coverage, the effective treatment group might coincide with the entire sample.

3.5 Results

In this section we report results for our analyses from the cross-sectional analyses, the map exposure experiment, and the longitudinal study.

3.5.1 Cross sectional analysis

We first perform a cross-sectional analysis of what factors might drive perceived flood risk, worry, and insurance willingness to pay. In this analysis we consider recent flooding experience. First, we consider the impact of Hurricane Harvey. After making landfall on August 25, 2017, Harvey battered parts of eastern Texas and Louisiana with more than 50 inches of rain, leading to catastrophic flooding that made international headlines (Sullivan, Hernandez and Fahrenthold 2017). We take respondents living within the boundaries of the Hurricane Harvey disaster area in Texas and Louisiana as our treatment group (regardless of whether the respondent personally experienced a flood). Additionally, we consider self reported flooding, defined as any flooding in the home in the past year anywhere in the region. The results are displayed in Table 3.4.

While recent flooding is not associated with greater flood risk perceptions, conditional on having a past flood in the home, residence in the Harvey disaster zone is significantly associated with greater perceived risk and more flood-related worry (Specifications 1-3). However, interpretation of these results as a Hurricane Harvey treatment effect is difficult without any responses pre-dating the disaster that could permit a difference-in-difference estimate.

In addition to recent storm exposure, we also consider how the response variables change in relationship with home elevation, walking distance to water, historical flood experience, and whether the respondent is a climate believer. We find that proximity to water, historical flooding experience, and being a climate behavior

Table 3.4 · Recent storm exposure and response variables

	Log risk (1)	Worry (2)	Ins. WTP (3)	Log risk (4)	Worry (5)	Ins. WTP (6)
Harvey disaster area	0.544**	0.906***	-70.88			
	(0.213)	(0.225)	(100.2)			
Home flooded in 2017 or 2018				0.277	-0.401	-265.7
				(0.327)	(0.379)	(305.6)
Home elevation (ft)	-0.00401**	-0.00650***	-2.006	-0.00512	-0.0116***	-3.151
	(0.00163)	(0.00156)	(1.428)	(0.00326)	(0.00267)	(2.174)
Lives walking dist. to water	0.740***	1.032***	67.48	0.717***	0.891***	-1.971
	(0.114)	(0.103)	(61.57)	(0.164)	(0.138)	(93.68)
County floods since 1965	-0.0351	-0.00802	15.73	1.138***	0.200	68.20
	(0.0272)	(0.0290)	(14.83)	(0.181)	(0.183)	(109.4)
Municipality floods since 1990	-0.0253	0.106***	7.853	-0.903***	1.341***	-108.0
	(0.0397)	(0.0385)	(20.45)	(0.318)	(0.393)	(256.0)
Past flood in home	1.478***	1.913***	545.3***	1.382***	1.948***	707.6**
	(0.108)	(0.158)	(122.4)	(0.235)	(0.278)	(270.1)
Climate believer	0.393***	0.624***	57.62	0.391**	0.652***	184.6*
	(0.120)	(0.140)	(84.96)	(0.187)	(0.197)	(97.42)
Demographic controls	✓	✓	✓	✓	✓	✓
Municipality fixed effects				✓	✓	✓
Observations	1,300	1,300	1,180	1,300	1,300	1,180
Adjusted R^2	0.160	0.214	0.051	0.185	0.213	0.129

OLS estimates with standard errors clustered at the county level in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Willingness to pay for flood insurance is expressed in terms of annual premia for a $100,000 policy. The first two independent variables listed are indicator variables. Respondents were considered to be in the Harvey disaster area if their home counties were included in the FEMA disaster declaration for permanent public assistance (i.e. not merely emergency assistance in the days after the flood). Demographic controls (not shown) comprise gender, homeowner status, tenure in current home, and categorical bins for age, race, income, educational attainment, home type, and risk preference. Specifications 4-6 include municipality fixed effects

are all positively correlated with risk, worry, and insurance willingness to pay. Further, we find that elevation is negatively correlated with risk and worry and that there is no significant relationship between elevation and willingness to pay for insurance.

3.5.2 Map exposure treatment

Figure 3.3 displays histograms showing the distribution of two key response variables across experimental ($n = 365$) and control ($n = 1{,}217$) groups. The left-hand figure tells a basic story: exposure to FEMA flood maps made the average respondent much more sanguine about flood risk, reducing perceived ten-year flood risk on average by 0.524 log points, or 41% ($p = 0.0003$). Respondents who estimated their risk at exactly 0% were imputed a value of $e^{(-5)} = 0.0067$. This effect is statistically and economically significant. We find that map exposure has a similarly strong and negative effect on self-reported worry about floods on a 1-10 scale (not shown).

Exposure did not significantly affect the spread of the distribution, but generally redistributed mass from the right side of the distribution, corresponding to flood risk of 1% per year or more, to the left side of the distribution. The qualitative responses of participants reinforce this finding, but also hint at significant heterogeneity in the effect across individuals. After examining the maps, 27% of respondents reported feeling their homes were less at risk, as compared to 14% who reported feeling more at risk.

However, the right-hand chart in Figure 3.3 shows the effect was only marginally significant in the case of willingness to pay for a $100,000 flood insurance contract, with exposure to the map depressing respondents' willingness to pay for insurance by $134 per year ($p = 0.07$).

Figure 3.4 presents a calibration exercise that analyzes the size of this effect in the context of the FEMA flood map risk estimates. The probability ranges are chosen to correspond to the three flood zones demarcated by the flood maps: the 100-year zone, the 500-year zone, and the region outside both zones, under the assumption that the maps are accurate and that each year's flood activity can be represented by independent Bernoulli draws. For example, a respondent living in the 100-year flood zone with annual risk of flooding equal to at least 0.01, would have a 10-year flood likelihood of:

$$p \geq 1-(1-0.01)^{10} = 0.0965.$$

Fully 26% of the sample reports living in the 100-year flood plain, as compared to an estimated 3% of the U.S. population as a whole (Crowell et al. 2010). The gray columns represent the implicit distribution of ten-year flood risk under these assumptions. Thus, the gray column in the third group on the right of the figure represents 26%.

The blue and red columns, by contrast, represent the actual answers provided by respondents, grouped into the same three ranges. For example, approximately 15% of the control group and 10% of the experimental group report their ten-year flood risk is greater than 9.65%. While there is no assurance that the distribution of the control group across flood zones matches the experimental group's, the fact that the groups are randomly assigned and very well-balanced on observable characteristics (Table 3.1) suggests this is a valid comparison.

Under the assumption that approximately 26% of the control sample is also in the 100-year flood zone, we find that the control group is more optimistic about flood risk than the FEMA maps, with only 15% reporting the high ten-year flood risk that is implied by the definition 100-year flood plain. This could be explained a number of ways aside from computational error: perhaps the participants do not think flooding across years is serially uncorrelated, do not think the flood map averages apply to them if they have taken special measures to protect their homes. It could also be the case that they do not take the flood maps at face value, although it should be noted that critics believe the FEMA flood maps tend to err on the side of understating risk (Kelly 2017).

Nevertheless, exposure to flood maps does not pull respondents closer to the distribution given by the gray columns; in fact, it pushes them further away. As shown above, exposure to flood maps has a negative impact on flood risk perceptions; people become more optimistic about flood risk. The differential distributions of the blue and red columns in this figure accords with that finding, and is highly statistically significant according to a χ^2 test of homogeneity.

One potential explanation is that the disclosure of information tends to make people feel more secure and less vulnerable, even if the information itself is adverse (Schade, Kunreuther and Koellinger 2012). Perhaps respondents' confidence in government disaster preparations is inspired by the very existence of the detailed FEMA flood maps. That could explain the especially strong effect on self-reported worry, which might capture this security-blanket effect.

Table 3.5 displays regression results for OLS models estimating the average map exposure effect on various response variables. Specifications 3, 6, and 9 in the table match Equation (2) above; the others are pared-down specifications that exclude hydrological controls, demographic controls, and municipality fixed effects.

The first set of specifications concern the effect of map exposure on perceived flood risk. Specification 1 corresponds most closely to Figure 3.3 above, and shows the large and statistically significant effect of map exposure on beliefs. This effect is robust to inclusion of controls and municipality fixed effects. Specification 2 demonstrates that home elevation is negatively associated with perceived flood risk, and that proximity to the water, past flooding in one's home, and belief that climate change is occurring are positively associated. Similar associations are found for flood worry, but generally not for willingness to pay, except for past flood experience and home elevation.

County flood history, a fairly imprecise proxy for local hydrological conditions, is not significantly associated with beliefs, although it is positively associated with self-reported worry. Additional results (shown in Appendix Table A3) using municipality-level historical data were not notably different from results shown here, and restricted the sample due to data availability issues.

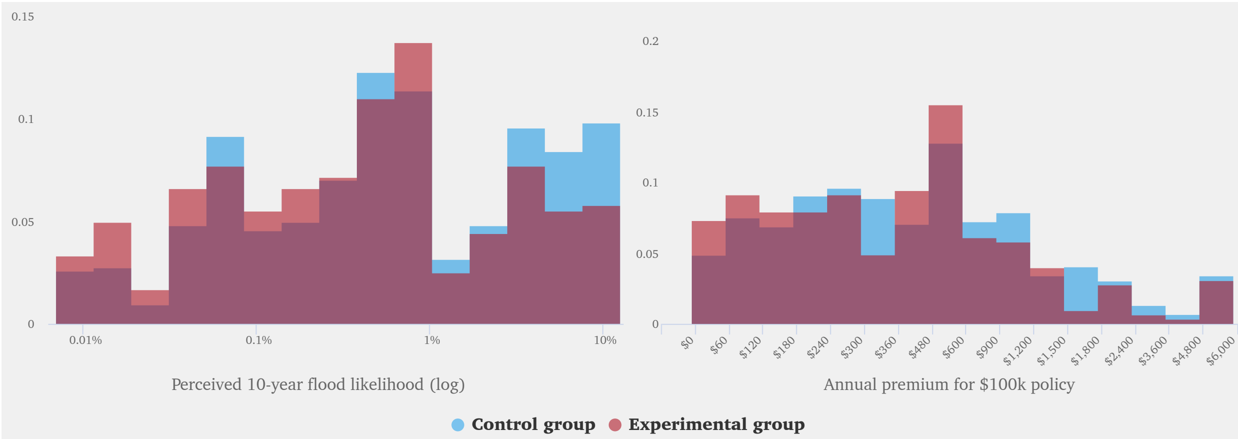


Figure 3.3 · Effect of exposure to FEMA flood maps

The histograms above show the distribution of two quantities, flood beliefs (left) and willingness to pay for flood insurance (right), separately for the control group and the experimental group. The x-axis on the left-hand chart is displayed on a log scale as the distribution is roughly lognormal, and the analysis that follows focuses on the logged version of this variable. The x-axis on the right-hand chart is nonlinear and reflects the 16 choices elicited on the multiple price list question (see Appendix A6, Question 29).

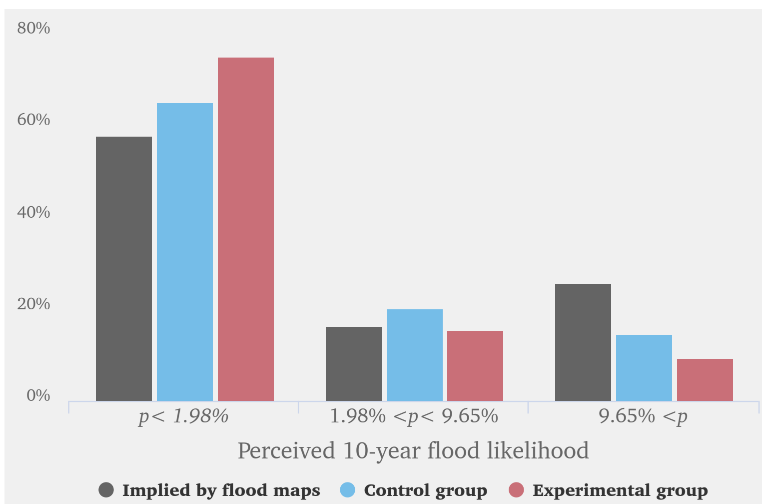


Figure 3.4 · Calibration of flood risk perceptions to FEMA flood maps

The distribution of perceived 10-year flood likelihoods across three intervals. The distribution of implicit flood probabilities (gray bars) are obtained by assuming the risk of annual flooding implied by the FEMA flood maps is correct, and that flood outcomes are independent across years. For example, 26% of respondents reported living in the 100-year flood zone, implying a 10-year flood risk of 9.65% or greater for 26% of the sample. The control group (blue) and experimental group (red) distributions, by contrast, are based on actual responses to the survey question about perceived flood likelihood.

Table 3.5 · Effect of flood map exposure

	Perceived 10-year flood likelihood (log)			Worry about floods (1-10 scale)			WTP for flood insurance		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Map exposure	-0.524***	-0.470***	-0.589***	-0.674***	-0.546***	-0.705***	-134.3	-93.73	-10.70
	(0.146)	(0.141)	(0.208)	(0.130)	(0.120)	(0.190)	(86.04)	(84.99)	(136.1)
Home elevation (ft)		-0.006***	-0.004		-0.011***	-0.011***		-2.161**	-3.061
		(0.002)	(0.004)		(0.002)	(0.003)		(1.072)	(2.116)
Live walking dist. to water		0.759***	0.774***		1.040***	0.951***		64.31	17.42
		(0.104)	(0.172)		(0.101)	(0.138)		(53.63)	(88.99)
County floods since 1965		-0.00677	0.647***		0.058	1.163***		17.29	-16.94
		(0.0265)	(0.180)		(0.0379)	(0.174)		(11.56)	(101.6)
Past flood in home		1.602***	1.472***		2.084***	1.687***		483.6***	500.3***
		(0.113)	(0.137)		(0.143)	(0.210)		(106.8)	(170.7)
Climate believer		0.376***	0.412**		0.592***	0.637***		19.18	123.6
		(0.111)	(0.178)		(0.125)	(0.197)		(81.65)	(116.4)
Hydrological controls		✓	✓		✓	✓		✓	✓
Demographic controls		✓	✓		✓	✓		✓	✓
Municipality fixed effects			✓			✓			✓
Observations	1,582	1,565	1,565	1,582	1,565	1,565	1,429	1,417	1,417
Adjusted R^2	0.010	0.174	0.199	0.0122	0.226	0.255	0.002	0.045	0.094

OLS estimates with standard errors clustered at the county level in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Hydrological controls (shown) comprise home elevation, proximity to the water, county-level historical flooding data, information on past floods in the home, and climate change beliefs. Demographic controls (not shown) comprise gender, homeowner status, tenure in current home, and categorical bins for age, race, income, educational attainment, home type, and risk preference. Specifications 3, 6, and 9 include 551 municipality fixed effects. A small number of observations with missing or unreliable data for demographic controls and are omitted from some regressions; see Table 3.3.

Also notable is the extent to which including municipality fixed effects mutes the association between home elevation and the response variables (compare Specifications 2 and 3, and 8 and 9). This is most likely due to the fact that many respondents (21%) are situated in municipalities that only appear once in the data, and the median observation is in a municipality with only five observations. Evidently, the fixed effects are soaking up most of the variation in elevation between respondents. Any residual, within-municipality variation in elevation, conditional on proximity to the water, does not seem to explain beliefs about flooding or willingness to pay for insurance. The imputation process for elevation discussed in Section 3.3.5 above could also be artificially reducing within-municipality variation in elevation. Below in Section 3.5.4, we perform a robustness check where observations with imputed elevation data are discarded and do not find significant changes.

In the preceding section, we argued that a regression model with interaction terms could be used to analyze exposure to flood maps as an information treatment. Table 3.6 displays specifications with interaction terms as detailed in Equation (3). In each set of specifications, the first specification uses only county-level flood history, while the second uses city-level flood history, at the expense of excluding some observations for which no municipality data exists. The results are broadly similar, so we focus on the county-only specifications.

The lower panel of Specification 1 labelled “Interaction effects” shows that flood beliefs become more closely correlated with home elevation when respondents are exposed to flood maps. The interaction effect is nearly double the size of the main effect, suggesting that home elevation becomes roughly three times as important in explaining perceived flood risk once respondents are exposed to the maps.

A belief that climate change is happening, which is strongly associated with stated flood beliefs and self-reported worry, becomes much less of a factor after map exposure. In Specification 1, the combined effect is a precisely estimated zero; climate change attitudes are no longer a powerful explanatory factor in predicting beliefs when respondents are exposed to the concrete flood information.

Similar results hold in Specification 3, but are marginally significant, and the map exposure treatment has no

Table 3.6 · Interaction between map exposure and hydrological characteristics

	Perceived flood likelihood (log)		Worry about floods		WTP for flood insurance	
	(1)	(2)	(3)	(4)	(5)	(6)
Main effects						
Home elevation (ft)	-0.00355*	-0.00285	-0.00900***	-0.00649***	-1.715	-1.578
	(0.00205)	(0.00174)	(0.00177)	(0.00187)	(1.163)	(1.537)
Lives walking dist. to water	0.743***	0.691***	0.988***	0.978***	60.71	62.60
	(0.118)	(0.128)	(0.104)	(0.107)	(63.95)	(69.31)
Municipality floods since 1990		-0.0237		0.103*		-0.153
		(0.0464)		(0.0529)		(22.24)
County floods since 1965	-0.0194	-0.0183	0.0561	0.0383	20.87	20.14
	(0.0245)	(0.0254)	(0.0348)	(0.0361)	(14.22)	(16.23)
Past flood in current home	1.659***	1.571***	2.136***	2.052***	500.0***	531.0***
	(0.124)	(0.130)	(0.169)	(0.181)	(132.6)	(142.8)
Climate believer	0.488***	0.498***	0.697***	0.741***	3.818	27.00
	(0.130)	(0.138)	(0.135)	(0.150)	(85.95)	(91.24)
Interaction effects						
Home elevation * Map	-0.0131***	-0.0146***	-0.0106***	-0.0106*	-2.708	-2.750
	(0.00399)	(0.00498)	(0.00360)	(0.00550)	(2.423)	(3.366)
Lives walking distance * Map	0.0173	0.0148	0.185	-0.0222	-15.16	-10.48
	(0.229)	(0.263)	(0.247)	(0.276)	(133.6)	(156.8)
County floods * Map	0.0634	0.0498	0.0146	-0.0143	-13.70	-35.28
	(0.0496)	(0.0610)	(0.0409)	(0.0503)	(28.63)	(27.58)
Municipality floods * Map		0.0136		0.102		25.28
		(0.0956)		(0.0822)		(44.78)
Past flood * Map	-0.329	-0.269	-0.302	-0.348	-71.27	39.00
	(0.311)	(0.346)	(0.311)	(0.328)	(355.0)	(385.9)
Climate believer * Map	-0.488*	-0.434*	-0.481**	-0.466*	54.83	133.7
	(0.248)	(0.259)	(0.205)	(0.238)	(137.6)	(143.7)
Demographic controls	✓	✓	✓	✓	✓	✓
Municipality-level data		✓		✓		✓
Observations	1,565	1,300	1,565	1,300	1,417	1,180
Adjusted R^2	0.180	0.168	0.229	0.213	0.0434	0.0494

OLS estimates with standard errors clustered at the county level in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Perceived flood likelihood is the respondent's stated belief that her home will flood in the next ten years. Worry about floods is expressed on a 1-10 scale, with greater numbers representing more worry. Willingness to pay for flood insurance is expressed in terms of annual premia for a$100,000 policy. Demographic controls (not shown) comprise gender, homeowner status, tenure in current home, and categorical bins for age, race, income, educational attainment, home type, and risk preference. Some specifications include municipality-level flood history from FEMA, which is not available for some respondents because their municipalities do not appear in the FEMA database. Municipality fixed effects are not included in these specifications; see text for further discussion.

significant effect on how hydrological factors contribute to perceptions of cost or willingness to pay for insurance.

The results in Specification 1 suggest important heterogeneity masked in Figure 3.3. People with homes at low elevations become relatively more pessimistic when exposed to the maps, while people who believe that climate change is occurring became relatively more positive.

The fact that the "learning" interaction effects point in the expected direction of the *main* effects is important. The effect of map exposure is apparently similar to a reminder that home elevation is important to determining one's flood risk. This is not literally the mechanism; indeed, respondents may not know their homes' elevations as they answer the question. But evidently the information contained in the map works to simulate this sort of hydrological understanding.

The apparent disconnect between the results for beliefs about future flooding and actual demand for insurance against that flooding is discussed more in the next section.

3.5.3 Hurricane exposure treatment

After the 2018 hurricane season, we re-sample respondents both inside and outside the afflicted area and perform the difference-in-difference analysis outlined in Section 3.4.3 and Equation 4. The two primary treatment regions we consider in this analysis are those that were effected by Hurricane Harvey in 2017 and Hurricane Florence in 2018. We present these treated regions in Figure 3.5.

We present the results for the counties that were specifically affected by Hurricane Florence, as defined by the declared disaster areas. These results are presented in Table 3.7. Overall, both risk and worry increase after the 2018 hurricane season across the sample. No significant effect is found on insurance willingness to pay. Additionally, we show that there is no significant differential response between waves in the Florence region specifically. One potential explanation for why we do not see a significant change in any of the variables of interest is because the number of respondents in our treatment region is relatively small (only 40 responses). A second potential reason is that even within those 40 responses, there may be heterogeneity in the type of flooding the respondents experienced. Thus in our next treatment we consider those who reported actually experiencing flooding between waves, regardless of their location.

In Table 3.8, we present the results from our differences in differences analysis, where our treated respondents are those that self-reported flooding in their homes. We find that after experiencing flooding in between the two waves, these respondents see an increase in perceived risk significant at the 10 percent level, but no change in either self-reported worry nor insurance willingness to pay.

Finally, we consider our third treatment region: those who experienced flooding prior to our initial wave, but not in 2018 ($n = 52$). We focus on this group to see how their responses change as time increases from the most recent flooding experience. Because our initial survey is close to their flooding experience, we expect that their responses may attenuate slightly in the second wave. This hypothesis is due to prior literature findings on risk salience. See Gallagher (2014). Interestingly, we find strong evidence that both risk and worry decrease significantly in the second wave for our region of interest relative to respondents overall (Table 3.9). This quick attenuation in worry and risk perception is consistent with the findings in the literature of relatively abrupt increases and decreases in insurance demand. Our coefficient on willingness to pay for insurance is negative, but not significant.

3.5.4 Robustness checks

To assess the robustness of our findings, we perform several checks using alternative specifications, data sources, and sample restrictions. In general, the main results – the average map exposure effect and the differential map exposure effect by home elevation – hold under these alternative modeling choices.

The first alternative we explore is including municipal flood history at the expense of excluding certain observations without data. As discussed in Section 3.3.2, municipal-level data on flood history is available from 1990 to 2007 from Gallagher (2014), but only covers approximately 82% of the sample. The remaining 18% of observations include respondents living in rural areas not covered by any formal municipality, or who list municipality names that do not correspond to any municipality listed in the data. While excluding these data will reduce statistical power, it allows us to control for local flood history on a much more granular level.

Specifications 2, 4, and 6 in Table 3.6 include municipality-level data, and consequentially the sample is restricted in those models. The results are largely robust in that table, but we also rerun the models from Table 3.5 to confirm that including municipality data does not affect those findings. The results are quite similar and are shown in Table A3.

Figure 3.5 · Natural experiment: hurricane exposure

This figure displays the tracks of two recent hurricanes. Any counties that were i) designated as disaster areas after each storm and ii) included in our survey sample are highlighted. Other counties included in our survey sample are denoted in dark gray.

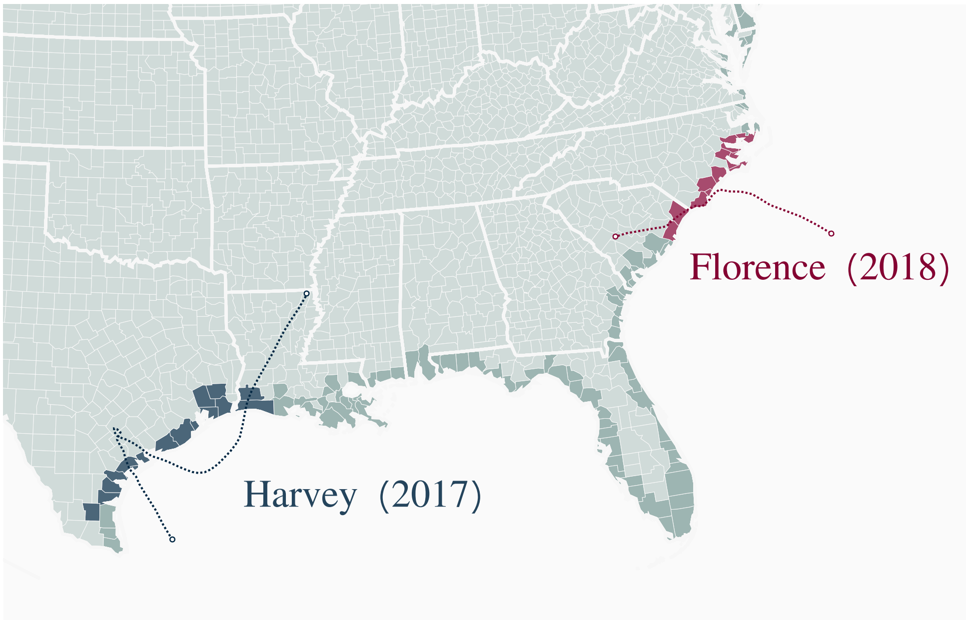


Table 3.7 · Treatment 1: Florence impact zone

	(1)	(2)	(3)
	Log risk	Worry (1-10)	Ins. WTP
Post	0.200**	0.314***	30.27
	(0.0811)	(0.0986)	(48.69)
Post × Florence zone	0.270	-0.167	-77.90
	(0.209)	(0.254)	(126.2)
Respondent fixed effects	✓	✓	✓
Constant	-0.573***	4.273***	768.6***
	(0.0522)	(0.0642)	(31.50)
Observations	1,426	1,448	1,360
R-squared	0.791	0.774	0.757

OLS estimates with standard errors clustered at respondent level in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The regressions are on the respondent-wave level, with each respondent represented by two observations, one in the first wave and then one in the second or third wave (whichever is the last response from that individual). The "Post × Florence zone" indicator is 1 for observations that are from the second or third wave, and for respondents in the Florence disaster area. A small number of observations with missing or unreliable data for demographic controls and are omitted from some regressions; see Table 3.3.

Table 3.8 · Treatment 2: Flooded between waves

	(1)	(2)	(3)
	Log risk	Worry (1-10)	Ins. WTP
Post	0.207***	0.276***	8.818
	(0.0765)	(0.0932)	(45.81)
Post × Flood between waves	0.663*	0.267	247.3
	(0.342)	(0.424)	(229.4)
Respondent fixed effects	✓	✓	✓
Constant	-0.573***	4.273***	768.6***
	(0.0521)	(0.0642)	(31.49)
Observations	1,426	1,448	1,360
R-squared	0.792	0.774	0.757

OLS estimates with standard errors clustered at respondent level in parentheses, *** p<0.01, ** p<0.05, * p<0.1. These specifications are analogous to those shown in Table 3.7, but where the "Flood between waves" indicator is 1 for respondents who report their home is flooded between the first wave and the last wave of the survey they complete. The regressions are on the respondent-wave level, with each respondent represented by two observations, one in the first wave and then one in the second or third wave (whichever is the last response from that individual). A small number of observations with missing or unreliable data for demographic controls and are omitted from some regressions; see Table 3.3.

Table 3.9 · Treatment 3: Exposed in '16 or '17, not '18

	(1)	(2)	(3)
	Log risk	Worry (1-10)	Ins. WTP
Post	0.253***	0.349***	23.74
	(0.0791)	(0.0951)	(46.65)
Post × Flood in '16/'17	-0.672**	-1.127***	-252.4
	(0.304)	(0.372)	(191.9)
Respondent fixed effects	✓	✓	✓
Constant	-0.607***	4.225***	757.7***
	(0.0533)	(0.0650)	(31.81)
Observations	1,356	1,378	1,300
R-squared	0.794	0.778	0.756

OLS estimates with standard errors clustered at respondent level in parentheses, *** p<0.01, ** p<0.05, * p<0.1. These specifications are analogous to those shown in Table 3.7, but where the "Flood in '16/'17" indicator is 1 for respondents who report a flood in one of those years, but in 2018 between the waves. The regressions are on the respondent-wave level, with each respondent represented by two observations, one in the first wave and then one in the second or third wave (whichever is the last response from that individual). A small number of observations with missing or unreliable data for demographic controls and are omitted from some regressions; see Table 3.3.

Our second alternative modeling choice is excluding respondents who failed attention check questions from our survey. We include two attention questions to test whether respondents are rushing through the survey without carefully considering the questions or their responses (see Appendix A6, Questions 26 and 22). A relatively large number of respondents failed at least one attention question, indicating that many respondents were not fully invested in answering the questions with care. This might simply add noise to the data, but it could also create artificial patterns or correlations if certain respondents always select the first answer in a multiple choice list, for example. However, throwing out responses from participants who failed the attention screen will significantly restrict the sample size, reducing statistical power, and may be especially compromising if the hurricane we will eventually use for our flood exposure quasi-experiment hits an area with relatively few respondents.

To assess the seriousness of this problem, we rerun the models displayed in Tables 3.5 and 3.6 without any of the respondents who failed either attention question. The results are shown in Appendix Tables A4 (main effect) and A5 (interaction effect), and are quite similar with the full-sample results. This indicates that the inclusion of the faulty respondents did not skew the main results.

Finally, we test versions of our model where respondents with imputed home elevations are excluded. Each respondent was asked to provide her home's elevation in feet, either from memory or by entering her address into a third-party service that provides elevation information (See Appendix A6, Question 32). Several of the provided answers were deemed unreliable because they exceeded the maximum elevation of their respective counties according to recreational mountain climbing website `www.peakbagger.com`. In cases where listed home elevations exceeded county high points, we imputed elevation by first using the Google Earth API to find latitude and longitude coordinates for each respondent's ZIP code, and then finding the elevation at that point.

Because elevation can vary significantly even within a single ZIP code, and because even a difference of a few feet in elevation can make a dramatic difference in flood risk near the shore, this imputation process potentially introduces significant measurement error. To assess this possibility, we rerun the models displayed in Tables 3.5 and 3.6 without any of the observations affected by the imputation procedure. The results are shown in Appendix Tables A6 (main effect) and A7 (interaction effect).

The results for the interaction effect are similar to the results from Table 3.6, and there is a significant negative association between willingness to pay for insurance and home elevation in Specification 3, as well as a significant positive association between WTP and county flood history, both of which are absent from the analogous Specification 8 in Table 3.5. These coefficients have the expected sign and are statistically significant, which suggests that the imputation process may be adding more noise than it is worth.

3.6 Conclusion

Increasing disaster risk driven by climate change highlights the need to understand how respondents update beliefs about disaster risk and pursue adaptation strategies such as insurance purchase. We analyze the results of a longitudinal survey of 124 coastal U.S. counties fielded in two waves, before and after the 2018 hurricane season, to answer this question in the context of home flooding. In order to focus on those who are most likely to be affected by hurricanes, we include coastal counties on the Atlantic seaboard and the Gulf Coast of Mexico in our survey region. The first wave of our survey was fielded on the Amazon mTurk survey platform during June and July 2018, prior to the bulk of the 2018 hurricane season, and yielded 1,582 responses. After the 2018 hurricane seasons ended, we sent follow-up surveys to our initial respondents. This second wave of surveys ran during the first half of 2019 and yielded 725 responses.

We begin by analyzing cross-sectional data from the first wave of our survey to understand the main factors that contribute to perceived risk, willingness to pay for hypothetical insurance, and actual flood insurance take-up. We find that proximity to water has a positive and significant relationship with all three dependent variables, and home elevation has a negative and significant relationship with all three variables, in line with a basic rational expectations model of insurance demand. We also find that being younger, a self-identified climate-change believer, and more risk averse are associated with greater perceived risk and willingness to pay for insurance.

In the first wave of our survey, we also conduct a randomized experiment to test how coastal residents react to government-provided flood maps. A subsample was randomly selected to visit the online flood map portal maintained by the Federal Emergency Management Agency (FEMA), find the flood maps for their neighborhoods, and receive a very brief tutorial on reading the maps. Exposure to these maps significantly reduced risk perceptions on average, but did not affect willingness to pay.

Next, we exploit the longitudinal nature of our data to measure how hurricane shocks affect perceived risk, willingness to pay, and insurance take-up over time. Given heavy hurricane activity in 2017 relative to 2018, we focus on attenuation of risk perceptions as disasters recedes from memory, especially in areas like coastal Texas that were battered by Hurricane Harvey in 2017 but saw no follow-up hurricanes in 2018. We find evidence that people whose homes were flooded in 2016 or 2017 – but not in 2018 between the first and second waves of the survey – report significantly higher perceived risk in the initial wave than the follow-up wave (50-60% average decline). We do not find that residents who were shocked by local floods in 2018 between survey waves, such as those in North Carolina affected by Hurricane Florence, exhibit increased risk or increased willingness to pay from one wave to the next, but the test is under-powered due to low sample size.

This work contributes to the literature in four main ways. First, we are able to confirm that basic risk factors like home elevation and proximity to the water can partially explain respondents' disaster beliefs and insurance demand. Second, we show that that climate change beliefs are correlated with perceived risk and willingness to pay for insurance at the individual level. Third, we are able to test the effect of exposure to flood information (separate from flood damages) utilizing FEMA's flood maps and a randomized experiment. And finally, we use novel longitudinal microdata to provide evidence of a mechanism – dramatic belief updating – that can explain the puzzling surges of insurance demand observed in prior work.

These findings have implications for how policymakers should try to help consumers process changes in risk, which will be important as flood risk changes rapidly in low-lying areas prone to sea level rise. The cross-sectional evidence suggests that people understand that proximity to water increases the likelihood of flooding and high elevation decreases chances of flooding, and that climate change is associated with an increased risk of flooding. These findings highlight to policy-makers what constituents already understand and demonstrate the potential value of educating constituents on climate change.

Our findings from the longitudinal study and the map experiment also provide behavioral insights that should be incorporated into information interventions pertaining to flood risk. As flooding risk changes rapidly in response to climate change, exposure to sporadic flooding may not cause the smooth, persistent belief updating that a simple economic model would predict. The juxtaposition between respondents' apparent ability to process risks associated with proximity and elevation, and the large swings in beliefs after hurricanes we see in the longitudinal data, suggests a complicated decision-making process that is not well-suited to climate adaptation.

Bibliography

Albouy, David, Walter Graf, Ryan Kellogg, and Hendrik Wolff. 2016. "Climate amenities, climate change, and American quality of life." *Journal of the Association of Environmental and Resource Economists*, 3(1): 205–246.

Atreya, A, and J Czajkowski. 2014. "Is flood risk universally sufficient to offset the strong desire to live near the water?" *Risk Management and Decision Processes Center, The Wharton School of the University of Pennsylvania*.

Baillon, Aurélien, Han Bleichrodt, and Vitalie Spinu. 2020. "Searching for the Reference Point." *Management Science*, 66: 93–112.

Baker, Andrew C, David F Larcker, and Charles CY Wang. 2022. "How much should we trust staggered difference-in-differences estimates?" *Journal of Financial Economics*, 144(2): 370–395.

Bakkensen, Laura A, and Lala Ma. 2020. "Sorting over flood risk and implications for policy reform." *Journal of Environmental Economics and Management*, 104: 102362.

Bakkensen, Laura A., and Lint Barrage. 2017. "Flood Risk Belief Heterogeneity and Coastal Home Price Dynamics: Going Under Water?" National Bureau of Economic Research Working Paper 23854.

Baldauf, Markus, Lorenzo Garlappi, and Constantine Yannelis. 2018. "Does Climate Change Affect Real Estate Prices? Only If You Believe in it."

Barseghyan, Levon, Francesca Molinari, Ted O'Donoghue, and Joshua C. Teitelbaum. 2018. "Estimating risk preferences in the field."

Bayer, Patrick, and Christopher Timmins. 2007. "Estimating equilibrium models of sorting across locations." *The Economic Journal*, 117(518): 353–374.

Bchir, Mohamed Ali, and Marc Willinger. 2013. "Does the exposure to natural hazards affect risk and time preferences? Some insights from a field experiment in Peru." LAMETA, Universitiy of Montpellier Working Papers.

Beetsma, Roel M.W.J., and Peter C. Schotman. 2001. "Measuring Risk Attitudes in a Natural Experiment: Data from the Television Game Show Lingo." *The Economic Journal*, 111: 821–848.

Bernheim, B Douglas, and Antonio Rangel. 2009. "Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics." *The Quarterly Journal of Economics*, 124(1): 51–104.

Bernstein, Asaf, Matthew T Gustafson, and Ryan Lewis. 2019. "Disaster on the horizon: the price effect of sea level rise." *Journal of Financial Economics*.

Bhargava, Saurabh, George Loewenstein, and Justin Sydnor. 2017. "CHOOSE TO LOSE: HEALTH PLAN CHOICES FROM A MENU WITH DOMINATED OPTIONS *." *The Quarterly Journal of Economics*, 1319–1372.

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

Blake, Eric S., Christopher W. Landsea, and Ethan J. Gibney. 2011. “The deadliest, costliest, and most intense United States tropical cyclones from 1851 to 2010.” National Hurricane Center.

Botzen, WJ Wouter, Howard Kunreuther, and Erwann Michel-Kerjan. 2015. “Divergence between individual perceptions and objective indicators of tail risks: Evidence from floodplain residents in New York City.” *Judgment and Decision Making*, 10(4): 365–385.

Boustan, Leah Platt, Matthew E Kahn, and Paul W Rhode. 2012. “Moving to higher ground: Migration response to natural disasters in the early twentieth century.” *American Economic Review*, 102(3): 238–44.

Browne, Mark J, and Robert E Hoyt. 2000. “The demand for flood insurance: empirical evidence.” *Journal of risk and uncertainty*, 20(3): 291–306.

Brown, Gardner M, and Henry O Pollakowski. 1977. “Economic valuation of shoreline.” *The Review of Economics and Statistics*, 272–278.

Bunten, Devin, and Matthew E Kahn. 2014. “The impact of emerging climate risks on urban real estate price dynamics.” National Bureau of Economic Research Working Paper 20018.

Cameron, Lisa, and Manisha Shah. 2015. “Risk-taking behavior in the wake of natural disasters.” *Journal of Human Resources*, 50(2): 484–515.

Cameron, Trudy Ann. 2005. “Updating Subjective Risks in the Presence of Conflicting Information: An Application to Climate Change.” *Journal of Risk and Uncertainty*, 30(1): 63–97.

Casey, Alexander. 2017. “12 Counties Account for a Third of U.S. Flood Insurance Claims.” Zillow.

Cassar, Alessandra, Andrew Healy, and Carl von Kessler. 2017. “Trust, Risk, and Time Preferences After a Natural Disaster: Experimental Evidence from Thailand.” *World Development*, 94: 90 – 105.

Champ, Patricia Ann, Geoffrey H. Donovan, and Christopher M. Barth. 2010. “Homebuyers and wildfire risk: A colorado springs case study.” *Society and Natural Resources*, 23(1): 58–70.

Charness, Gary, Uri Gneezy, and Alex Imas. 2013. “Experimental methods: Eliciting risk preferences.” *Journal of Economic Behavior & Organization*, 87: 43–51.

Chinco, Alex, and Christopher Mayer. 2016. “Misinformed speculators and mispricing in the housing market.” *The Review of Financial Studies*, 29(2): 486–522.

Crowell, Mark, Kevin Coulton, Cheryl Johnson, Jonathan Westcott, Doug Bellomo, Scott Edelman, and Emily Hirsch. 2010. “An estimate of the U.S. population living in 100-year coastal flood hazard areas.” *Journal of Coastal Research*, 201–211.

Davis, Lucas W. 2004. “The effect of health risk on housing values: Evidence from a cancer cluster.” *American Economic Review*, 94(5): 1693–1704.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken. 2014. “What Do We Learn from the Weather? The New Climate-Economy Literature.” *Journal of Economic Literature*, 52(3): 740–98.

Deryugina, Tatyana. 2013. “How do people update? The effects of local weather fluctuations on beliefs about global warming.” *Climatic Change*, 118(2): 397–416.

Deschênes, Olivier, and Michael Greenstone. 2007. “The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather.” *American Economic Review*, 97(1): 354–385.

Dessaint, Olivier, and Adrien Matray. 2017. “Do managers overreact to salient risks? Evidence from hurricane strikes.” *Journal of Financial Economics*, 126(1): 97–121.

Diamond, Alexis, and Jasjeet S Sekhon. 2013. "Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies." *Review of Economics and Statistics*, 95(3): 932–945.

Di Mauro, Carmela, and Anna Maffioletti. 2001. "The Valuation of Insurance under Uncertainty: Does Information about Probability Matter?" *The Geneva Papers on Risk and Insurance Theory*, 26(3): 195–224.

Eckel, Catherine, Mahmoud El-Gamal, and Rick K. Wilson. 2009. "Risk loving after the storm: A Bayesian-Network study of Hurricane Katrina evacuees." *Journal of Economic Behavior & Organization*, 69(2): 110–124.

Egan, Patrick J, and Megan Mullin. 2012. "Turning personal experience into political attitudes: The effect of local weather on Americans' perceptions about global warming." *The Journal of Politics*, 74(3): 796–809.

Federal Emergency Management Agency. 2016. "Summary of Disaster Declarations and Grants." *https://www.fema.gov/media-library-data/1493738442601-01db152481b5d3d747535ae0a1c441a6/DataVizDisasterSummariesFV12.19.2016.xlsx*.

Federal Emergency Management Agency. 2017. "National Flood Insurance Program Community Rating System Coordinator's Manual." *https://www.fema.gov/media-library-data/1493905477815-d794671adeed5beab6a6304d8ba0b207/633300*$_2$*017*$_C$*RS*$_C$*oordinators*$_M$*anual*$_5$*08.pdf*.

Federal Emergency Management Agency. 2019. "Disaster declarations summaries." *https://www.fema.gov/openfema-dataset-disaster-declarations-summaries-v1*.

Federal Emergency Management Agency. 2021. "National Flood Hazard Layer." *https://www.fema.gov/national-flood-hazard-layer-nfhl*.

First Street Foundation. 2020. "First Street Foundation Flood Model (FSF-FM) Technical Documentation." *https://assets.firststreet.org/uploads/2020/06/FSF*$_F$*lood*$_M$*odel*$_T$*echnical*$_D$*ocumentation.pdf*.

Florida FWC. 2021. "Florida Land Cover Classification System." *https://myfwc.com/research/gis/regional-projects/cooperative-land-cover/*.

Freeman, A Myrick, Joseph A Herriges, and Catherine L Kling. 2014. *The measurement of environmental and resource values: Theory and methods.* Routledge.

Gallagher, Justin. 2014. "Learning about an Infrequent Event: Evidence from Flood Insurance Take-Up in the United States." *American Economic Journal: Applied Economics*, 6(3).

Gibson, Matthew, Jamie T. Mullins, and Alison Hill. 2019. "Climate risk and beliefs: Evidence from New York floodplains." Department of Economics, Williams College Department of Economics Working Papers 2019-02.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics*, 225(2): 254–277.

Graff Zivin, Joshua S, Yanjun Liao, and Yann Panassie. 2020. "How hurricanes sweep up housing markets: Evidence from florida." National Bureau of Economic Research 27542.

Gul, Faruk. 1991. "A Theory of Disappointment Aversion." *Econometrica*, 59: 667–686.

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

Hanaoka, Chie, Hitoshi Shigeoka, and Yasutora Watanabe. 2018. "Do Risk Preferences Change? Evidence from the Great East Japan Earthquake." *American Economic Journal: Applied Economics*, 10(2): 298–330.

Hanemann, W Michael. 1982. "Applied welfare analysis with qualitative response models."

Hartley, Catherine A, and Elizabeth A Phelps. 2012. "Anxiety and decision-making." *Biological psychiatry*, 72(2): 113–118.

Hartley, Roger, Gauthier Lanot, and Ian Walker. 2014. "Who really wants to be a millionaire? Estimates of risk aversion from gameshow data." *Journal of Applied Econometrics*, 29(6): 861–879.

Hausman, J. A. 1978. "Specification Tests in Econometrics." *Econometrica*, 46(6): 1251–1271.

Heath, Chip, Richard P. Larrick, and George Wu. 1999. "Goals as Reference Points." *Cognitive Psychology*, 38: 79–109.

Hino, Miyuki, and Marshall Burke. 2021. "The effect of information about climate risk on property values." *Proceedings of the National Academy of Sciences*, 118(17).

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, 47(2): 263–292.

Kahn, Matthew E. 2016. "The climate change adaptation literature." *Review of Environmental Economics and Policy*, 10(1): 166–178.

Kaiser, Harry M, Susan J Riha, Daniel S Wilks, David G Rossiter, and Radha Sampath. 1993. "A farm-level analysis of economic and agronomic impacts of gradual climate warming." *American Journal of Agricultural Economics*, 75(2): 387–398.

Kamenica, Emir. 2008. "Contextual Inference in Markets: On the Informational Content of Product Lines." *American Economic Review*, 98(5): 2127–49.

Kelly, John V. 2017. "FEMA Needs to Improve Management of Its Flood Mapping Programs." Office of Inspector General, Department of Homeland Security, Washington, DC.

Kerns, Kristin, and L. Slagan Locklear. 2019. "Three New Census Bureau Products Show Domestic Migration at Regional, State, and County Levels. U.S. Census Bureau."

Keys, Benjamin J, and Philip Mulder. 2020. "Neglected no more: Housing markets, mortgage lending, and sea level rise." National Bureau of Economic Research 27930.

Konisky, David M, Llewelyn Hughes, and Charles H Kaylor. 2016. "Extreme weather events and climate change concern." *Climatic Change*, 134(4): 533–547.

Kousky, Carolyn. 2010. "Learning from extreme events: Risk perceptions after the flood." *Land Economics*, 86(3): 395–422.

Kousky, Carolyn. 2017. "Disasters as learning experiences or disasters as policy opportunities? Examining flood insurance purchases after hurricanes." *Risk analysis*, 37(3): 517–530.

Kusev, Petko, Harry Purser, Renata Heilman, Alex J Cooke, Paul Van Schaik, Victoria Baranova, Rose Martin, and Peter Ayton. 2017. "Understanding risky behavior: the influence of cognitive, emotional and hormonal factors on decision-making under risk." *Frontiers in psychology*, 8: 102.

Köbberling, Veronika, and Peter P. Wakker. 2005. "An index of loss aversion." *Journal of Economic Theory*, 122: 119–131.

Kőszegi, B., and M. Rabin. 2006. "A Model of Reference-Dependent Preferences." *The Quarterly Journal of Economics*, 121: 1133–1165.

Kőszegi, Botond, and Matthew Rabin. 2007. "Reference-Dependent Risk Attitudes." 97: 1047–1073.

Lansford, Node H., and Lonnie L. Jones. 1995. "Recreational and Aesthetic Value of Water Using Hedonic Price Analysis." *Journal of Agricultural and Resource Economics*, 20(2): 341–355.

Leggett, Christopher G. 2002. "Environmental valuation with imperfect information the case of the random utility model." *Environmental and Resource Economics*, 23(3): 343–355.

Loewenstein, George, and Jennifer S Lerner. 2003. "The role of affect in decision making."

Loomes, Graham, and Robert Sugden. 1986. "Disappointment and Dynamic Consistency in Choice under Uncertainty." *Source: The Review of Economic Studies*, 53: 271–282.

Loomes, Graham, and Robert Sugden. 1987. "Some implications of a more general form of regret theory." *Journal of Economic Theory*, 41(2): 270–287.

Loomis, John, and Marvin Feldman. 2003. "Estimating the benefits of maintaining adequate lake levels to homeowners using the hedonic property method." *Water Resources Research*, 39(9).

Mase, Amber Saylor, Benjamin M Gramig, and Linda Stalker Prokopy. 2017. "Climate change beliefs, risk perceptions, and adaptation behavior among Midwestern U.S. crop farmers." *Climate Risk Management*, 15: 8–17.

McClelland, Gary H, William D Schulze, and Don L Coursey. 1993. "Insurance for low-probability hazards: A bimodal response to unlikely events." *Journal of Risk and Uncertainty*, 7(1): 95–116.

McCoy, Shawn J, and Randall P Walsh. 2018. "Wildfire risk, salience & housing demand." *Journal of Environmental Economics and Management*, 91: 203–228.

McCoy, Shawn J., and Xiaoxi Zhao. 2018. "A City under Water: A Geospatial Analysis of Storm Damage, Changing Risk Perceptions, and Investment in Residential Housing." *Journal of the Association of Environmental and Resource Economists*, 5(2): 301–330.

Mebane Jr, Walter R, and Jasjeet S Sekhon. 2011. "Genetic optimization using derivatives: the rgenoud package for R." *Journal of Statistical Software*, 42: 1–26.

Mendelsohn, Robert, William D Nordhaus, and Daigee Shaw. 1994. "The impact of global warming on agriculture: a Ricardian analysis." *The American Economic Review*, 753–771.

Milon, J Walter, Jonathan Gressel, and David Mulkey. 1984. "Hedonic amenity valuation and functional form specification." *Land Economics*, 60(4).

MIT Election Data and Science Lab. 2017. "U.S. President 1976–2020." *https://doi.org/10.7910/DVN/42MVDX*.

Miu, Andrei C, Mircea Miclea, and Daniel Houser. 2008. "Anxiety and decision-making: Toward a neuroeconomics perspective." *Neuroeconomics*.

Mulder, Philip. 2021. "Mismeasuring Risk: The Welfare Effects of Flood Risk Information." Working Paper.

Muller, Nicholas Z, and Caroline A Hopkins. 2019. "Hurricane Katrina Floods New Jersey: The Role of Information in the Market Response to Flood Risk." National Bureau of Economic Research Working Paper 25984.

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

National Oceanic and Atmospheric Administration. 2021. "Sea Level Rise Viewer." *https://coast.noaa.gov/digitalcoast/tools/slr.html*.

Ortega, Francesc, and Suleyman Taspinar. 2018. "Rising sea levels and sinking property values: Hurricane Sandy and New York's housing market." *Journal of Urban Economics*, 106: 81–100.

Palmquist, Raymond B. 1992. "Valuing localized externalities." *Journal of Urban Economics*, 31(1): 59 – 68.

Palm, Risa. 1995. "The Roepke lecture in economic geography catastrophic earthquake insurance: Patterns of adoption." *Economic Geography*, 71(2): 119–131.

Paolacci, Gabriele, Jesse Chandler, and Panagiotis G Ipeirotis. 2010. "Running experiments on Amazon Mechanical Turk." *Judgment and Decision Making*, 5(5): 411–419.

Prelec, Drazen. 1998. "The probability weighting function." *Econometrica*, 497–527.

Rabin, Matthew, and Richard H Thaler. 2001. "Anomalies: risk aversion." *Journal of Economic perspectives*, 15(1): 219–232.

Reynaud, Arnaud, Cécile Aubert, and Manh-Hung Nguyen. 2013. "Living with Floods: Protective Behaviours and Risk Perception of Vietnamese Households." *The Geneva Papers on Risk and Insurance - Issues and Practice*, 38(3): 547–579.

Royal, Andrew, and Margaret Walls. 2019. "Flood Risk Perceptions and Insurance Choice: Do Decisions in the Floodplain Reflect Overoptimism?" *Risk Analysis*, 39(5): 1088–1104.

Said, Farah, Uzma Afzal, and Ginger Turner. 2015. "Risk taking and risk learning after a rare event: Evidence from a field experiment in Pakistan." *Journal of Economic Behavior & Organization*, 118: 167–183.

Sargan, John D. 1958. "The estimation of economic relationships using instrumental variables." *Econometrica: Journal of the Econometric Society*, 393–415.

Schade, Christian, Howard Kunreuther, and Philipp Koellinger. 2012. "Protecting Against Low-Probability Disasters: The Role of Worry." *Journal of Behavioral Decision Making*, 25(5): 534–543.

Shao, Wanyun, and Kirby Goidel. 2016. "Seeing is believing? An examination of perceptions of local weather conditions and climate change among residents in the US Gulf Coast." *Risk Analysis*, 36(11): 2136–2157.

Shr, Yau-Huo Jimmy, and Katherine Y Zipp. 2019. "The aftermath of flood zone remapping: The asymmetric impact of flood maps on housing prices." *Land Economics*, 95(2): 174–192.

Simmons, Ann M., and Les Neuhaus. 2017. "Tampa hasn't had a big hurricane in 96 years. That's about to change." *Los Angeles Times*.

Slovic, Paul. 2004. "What's fear got to do with it-It's affect we need to worry about." *Mo. L. Rev.*, 69: 971.

Slovic, Paul, Howard Kunreuther, and Gilbert F White. 1974. "Decision processes, rationality and adjustment to natural hazards." In . Earthscan Publications.

Snowberg, Erik, and Justin Wolfers. 2010. "Explaining the favorite–long shot bias: Is it risk-love or misperceptions?" *Journal of Political Economy*, 118(4): 723–746.

Stuart, Elizabeth A, Gary King, Kosuke Imai, and Daniel Ho. 2011. "MatchIt: nonparametric preprocessing for parametric causal inference." *Journal of statistical software*.

Sugden, Robert. 2003. "Reference-dependent subjective expected utility." *Journal of Economic Theory*, 111: 172–191.

Sullivan, Kevin, Arelis R. Hernandez, and David A. Fahrenthold. 2017. "At least 22 confirmed dead as Harvey pivots toward Louisiana." *Washington Post*.

Sydnor, Justin. 2010. "(Over)insuring Modest Risks." *American Economic Journal: Applied Economics*, 2: 177–199.

Taubinsky, Dmitry, and Alex Rees-Jones. 2018. "Attention variation and welfare: theory and evidence from a tax salience experiment." *The Review of Economic Studies*, 85(4): 2462–2496.

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

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

Tversky, Amos, and Daniel Kahneman. 1992. "Advances in prospect theory: Cumulative representation of uncertainty." *Journal of Risk and uncertainty*, 5(4): 297–323.

Union of Concerned Scientists. 2018. "Underwater: Rising seas, chronic floods, and the implications for U.S. coastal real estate. Union of Concerned Scientists."

University of Florida GeoPlan Center. 2019. "Florida Geographic Data Library." *https://download.fgdl.org/pub/state/coast_feb04.zip*.

Wibbenmeyer, Matthew, Sarah Anderson, and Andrew J Plantinga. 2016. "Risk Salience, Public Pressure, and Agency Action: Wildfire and the Management of Public Lands."

Yale Program on Climate Change Communication (YPCCC) George Mason University Center for Climate Change Communication (Mason 4C). 2020. "Climate Change in the American Mind: National survey data on public opinion (2008-2018)." *https://osf.io/jw79p/wiki/home/*.

Zhang, Lei, and Tammy Leonard. 2019. "Flood Hazards Impact on Neighborhood House Prices." *The Journal of Real Estate Finance and Economics*, 58(4): 656–674.

Zillow. 2021. "ZTRAX: Zillow Transaction and Assessor Dataset." *http://www.zillow.com/ztrax/*.

Appendices

A1 Observations excluded from analysis

I exclude transactions with any of the following characteristics:

- No sales price listed
- No date listed
- Missing data on property ZIP code or geographic coordinates
- Property location, address, or ZIP code not within Florida
- Missing or malformed assessor parcel number
- Sales price below $10,000
- Sales price above $10 million
- intra-family sale
- Multi-property sale
- Buyer is bank, religious organization, or government
- Cannot identify buyer due to duplicates in buyer name listings
- Geographic coordinates are outside listed county
- Missing or incomplete buyer mailing address
- Buyer address cannot be matched to any county

I also excluded some transactions based on assessment data linked to the property transacted. Transactions of properties with the following listed land uses are excluded:

- Mobile home
- Manufactured home
- Residential parking
- Residential common area
- Planned unit development
- Multi-family property
- Commercial property

An additional set of observations are included in the main analysis, but have data that is suspect or anomalous in some way, and are excluded in some analysis (see Table 1.5, Column 3). Observations were tagged as suspect if they had any of the following issues:

- More than two buyers listed
- Inconsistencies between the described land use and the listed number of units in a building
- Property is a vacant lot
- Contradictory indications from transaction data and assessment data about whether property is a condominium, or whether it was owner-occupied
- Flood zone listed as "D" (areas with indeterminate flood hazard)
- Inconsistency between NOAA sea level submersion zone and property elevation

A2 Same-address listing behavior

Same-address listing behavior varies significantly across counties, and across time in a predictable seasonal pattern. For my analysis, the relationship between same-address listing and disaster pressure (see Section 1.3.5) is of primary importance. To examine this, I estimate the following model using ordinary least squares:

$$s_{t,j} = \beta Q_{t,j} + \phi_{m(t)} + \xi_j + \epsilon, \tag{1}$$

where $Q_{t,j}$ and $s_{t,j}$ are the disaster pressure and same-address rate respectively in month t and Florida MSA j, ξ_j are MSA fixed effects, and $\phi_{m(t)}$ are calendar month fixed effects. I also estimate analogous MSA-level models of the form:

$$s_t = \beta Q_t + \phi_{m(t)} + \epsilon. \tag{2}$$

Figure A6 displays OLS estimates of these specifications. While in most cases the relationship is statistically significant and negative, the estimated values of β range between -0.5 and 0, suggesting that each percentage point of disaster pressure depresses same-address listing behavior by less than half a percentage point. Assuming same-address listing behavior is related to demographics, this suggests a compositional effect: disaster exposure is causing slightly more selection out of Florida real estate markets by same-address listers. This pattern alone cannot explain the retreat from the water, unless flood exposure differentially causes people buying near the water who would otherwise list origin addresses to list the same address.

A3 Identifying ZIP codes with physical flood damage

I use the following procedure to determine whether a particular ZIP code has been physically damaged by a flood disaster, for the purposes of the robustness check in Table 1.5, Column 2. The analysis is at the ZIP code-month level. I use public data from FEMA on the number of claims filed on National Flood Insurance Program (NFIP) policies in each ZIP code in each month. I consider a ZIP code z in month t to be shocked by disaster if there was a flood disaster in the county containing z in either month t or month $t-1$. This allows for the possibility of a storm striking late in the month leading to heightened claims in the following month. I then calculate the average number of claims in each ZIP code across months, conditional on no shock. I take this average to be the natural background rate of claims in each ZIP code that result from hyperlocal, idiosyncratic flooding events that would not plausibly treat the residents of that ZIP code to the same degree as a federally-declared disaster.

I then examine ZIP-months where the claim count was at least 10 times greater than the ZIP-specific background rate. This occurs less than 1% of the time in ZIP-months with no shock, but 21% of the time in ZIP-months with a shock. I use this criterion to determine whether a ZIP has sustained physical damage. I additionally include the few ZIP-months with more than 100 claims that do not otherwise qualify. For each ZIP code with a physical damage event, I consider all observations from that ZIP code for the following sixty months (5 years) to be tainted. This exclusion applies to 7.1% of all ZIP-months in the full sample, and 9% of the transactions.

One limitation of this approach is that I cannot reliably detect physical damage in ZIP codes with very low participation in NFIP. An area with very few policies in place might sustain direct physical damage without any apparent abnormalities in claiming behavior: claims counts will be very low over time and remain low during the disaster.

A4 Distinctive origin states

I determine each Florida county's most distinctive origin state using the following algorithm.

1. For each origin state-Florida county dyad, I calculate the share of non-local buyers in that county (that is, buyers from outside the local MSA) who originate from that state, using the full sample described in Table 1.1, Column 3. For example, 15.1% of known-origin non-local buyers in Duval County are from California, so the value for the California-Duval dyad is 15.1.

2. I rank the 1,470 dyads (49 states outside Florida × 30 Florida counties) in descending order of value.

3. I select the top dyad and designate the state in the dyad as the most distinctive origin state of the county in the dyad. The top-ranked dyad overall is Georgia-Bay, with a value of 34.8. This means 34.8% of known-origin non-local buyers in Bay County come from Georgia, a dominant share. Hence, Georgia is the most distinctive origin county for Bay County.

4. I eliminate all remaining dyads that include either the county or the state from the selected dyad. This prevents repeated assignments.

5. I repeat Steps 2 through 4 using the remaining list of dyads until all Florida counties have been assigned a state.

A5 Imputation of expected performance for online GoalQuest participants

We impute expected performance as follows. First, we calculate the difference in subjective likelihood of completing n and $n + 2$ grids, and assume the participant will complete exactly $n + 1$ grids with this likelihood. For example, if a participant reports a 60 percent likelihood of completing 8 grids and 90 percent likelihood of completing 6 grids, we impute an expectation they will complete exactly 7 grids with 20% likelihood. To address expectations about performance below 4 grids and above 18 grids, we take the observed average performance among participants who complete less than 4 grids and more than 18 grids. On average, these participants complete 1.09 and 19.89 grids respectively. If participants assign any likelihood to completing less than 4 grids or more than 18 (that is, if their subjective likelihood of completing 4 grids is less than 100%, or their subjective likelihood of completing 18 grids is greater than 0%), we impute these values as the expected conditional performance. Finally, we sum these expectations across the entire distribution for each participant to arrive at a total expected performance.

A6 Survey instrument

The questions included in the online survey from Chapter 3 are reproduced below. Some graphics and visual slider elements are omitted.

Demographics and location

1. Location. Enter your home municipality. The municipality is the town/city where you live.
2. Location. Enter your home ZIP code. You must enter a valid 5-digit U.S. ZIP code.
3. Gender. What is your gender?
4. Age. How old are you?
   - 18 - 24
   - 25 - 34
   - 35 - 44
   - 45 - 54
   - 55 - 64
   - 65 - 74
   - 75 - 84
   - 85 or older
5. Income. What is your total annual household income?
   - Less than $15,000
   - $15,000 - $29,999
   - $30,000-$44,999
   - $45,000-$59,999
   - $60,000-$74,999
   - $75,000-$99,999
   - $100,000-$149,999
   - $150,000-$199,999
   - Greater than or equal to $200,000
6. Ethnicity. What is your ethnicity or race?
   - Caucasian/White
   - Hispanic or Latino
   - African American
   - Asian/Pacific Islander
   - Native American/American Indian
   - Other
7. Education. What is the highest degree or level of schooling you have completed?
   - 8th grade or less
   - Some High School
   - High School Graduate or GED
   - Some College
   - Trade/Technical/Vocational Training
   - Associates Degree
   - Bachelor's Degree
   - Some Graduate School
   - Graduate Degree
8. Political Affiliation. What is your political affiliation?
   - Democrat
   - Independent
   - Republican
   - None
   - Other
9. Homeownership status. Which of the following best describes you?
   - I own my primary residence.
   - I do not own my primary residence, and pay rent to a landlord or property owner.
   - Other
10. Home type. What type of structure is your home?
    - Detached, single-family home
    - Townhouse or row house
    - Condo, apartment, or multiresidence
    - Mobile home
    - Other
11. How many years have you lived in your current residence?

Risk preferences

12. Risk attitudes. Suppose you must choose one of the following games to play. In each game, a fair coin is flipped, with a 50% chance of coming up heads and a 50% chance of coming up tails. You earn a prize that depends on which side comes up. Which of the following games would you most like to play?
    - Heads: you get $70 / Tails: you get $2
    - Heads: $60 / Tails: $12
    - Heads: $52 / Tails: $16

- Heads: $44 / Tails: $20
- Heads: $36 / Tails: $24
- Heads: $28 / Tails: $28

13. Gambling. Do you bet on lotteries, casinos, sporting events, or horse races?
    - Don't gamble at all
    - Used to gamble, but quit gambling now
    - Hardly gamble
    - Several times a year or so
    - Once a month or so
    - Once a week or so
    - Almost every day
14. Which of the following hypothetical payment options would you prefer? Indicate a preference for each pair of options.
    - A) $100 one month from now vs. B) $105 one year from now
    - A) $100 one month from now vs. B) $110 one year from now
    - A) $100 one month from now vs. B) $125 one year from now
    - A) $100 one month from now vs. B) $150 one year from now
15. How worried are you about the risk of an asteroid impact from outer space that destroys your city or town over the next 10 years, again on a scale of 1 to 10 where "1" means "not worried at all" and "10" means "very worried"? [slider, 1 - 10]

Situation of your home

16. How close do you live to a body of water that could flood your home? This would include the ocean, a lake, or a river.
    - Walking distance (could walk to the water in less than 20 minutes)
    - Driving distance
17. How long would it take you to walk to the closest body of water? [slider, 0 - 20 minutes]
18. How long would it take you to drive to the closest body of water? [slider, 0 - 120 minutes]
19. Do you think your home is at more or less risk of flooding than other homes nearby?
    - More risk
    - About the same
    - Less risk

Flood maps [this section assigned randomly to one quarter of the sample]

20. Please follow the steps below.
    - Access the FEMA Flood Map here [`https://msc.fema.gov/portal/search`]. The link will open in a new window or tab in your web browser so that the survey will not be interrupted.
    - Enter your exact street address with city and zip code in the box provided (see example below). Note: this information is being entered on an external website operated by the Federal Emergency Management Agency and will not be accessible to the researchers administering this study.
    - Find your home on the map. It is recommended that you zoom out to get a better view of your neighborhood.
    - See below for an example flood hazard map of downtown Newport, Rhode Island. Note the orange and blue zones near the coast. These zones indicate increased flood risk, and are more likely to appear near rivers, beaches, and other areas that are vulnerable to flood. Look to see if there are orange and blue zones on the map near your home. Your neighborhood may or may not have any orange or blue zones, depending on the flood risk in your area.
    - The blue zone on the flood hazard map is called the "100-year flood zone." The U.S. Army Corps of Engineers assesses flood risk at greater than 1% per year in this area. The blue areas on the map will be flooded approximately once every century. The orange zone on the flood hazard map is called the "500-year flood zone." The U.S. Army Corps of Engineers assesses flood risk at greater than 0.2% per year in this area. The orange areas on the map will be flooded approximately once every five centuries.
21. Flood risk. Which of the following is true of your home, according to the map?
    - My home is in the 100-year flood zone (blue zone)
    - My home is in the 500-year flood zone (orange zone)
    - My home is not in either flood zone, but other parts of my neighborhood are
    - I don't see any flood zones in my neighborhood
22. Which of the two color zones discussed on the previous page has a higher risk of flood?
    - Green zone
    - Blue zone
    - Red zone
    - Orange zone
23. Your reaction. Which of the following best describes your reaction to seeing the flood map for your area?
    - My home's flood risk is a lot higher than I thought
    - My home's flood risk is a bit higher than I thought
    - My home's flood risk is about what I thought
    - My home's flood risk is a bit less than I thought
    - My home's flood risk is a lot less than I thought

Flood insurance perceptions and demand

24. How worried are you about risk of your home flooding over the next 10 years, on a scale of 1 to 10? [slider, 1-10]
25. Chances of a flood in your home. Consider your residence and the surrounding area. What do you think are the chances of your home being flooded to a depth of at least one inch at any point in the next ten years? Note: do not consider internal water damage from e.g. leaky pipes. [After answering this question, respondents were presented with sliders that allowed them to pick an exact value.]
    - About one in a thousand (0.1%) or less
    - More than one in a thousand chance but less than one in a hundred chance
    - About one in a hundred (1%)

- More than one in a hundred but less than one in ten chance
- About one in ten (10%) or more

26. We asked you to consider the chances of your home flooding to at least what depth?
    - One inch
    - Six inches
    - One foot
    - Five feet
27. Cost of a flood. Now, imagine that your home floods to a depth of one foot. How much do you think it would cost to repair the damage to the building and replace belongings that are destroyed, in dollars? Consider how much it would cost in total, even if you couldn't afford to pay the entire amount or if some of the damage would be covered by insurance.
28. Of the total damages, about what percentage would you expect to be paid back through public assistance from the government? [slider, 0% - 100%]
29. Flood insurance. Assume you are offered a flood insurance policy that will cover you for the next year. It will cover any damages to your home caused by external flooding up to $100,000. Consider how much you would be willing to pay per month or per year for that insurance policy, compared to going without insurance. For each hypothetical premium below, indicate whether you think it would be worth it to buy flood insurance at that price or not. Note: if you are not a homeowner, imagine that you own the residence where you currently live and must make the purchase decision for that residence.
    - monthly premium $5 | annual premium $60
    - monthly premium $10 | annual premium $120
    - monthly premium $15 | annual premium $180
    - monthly premium $20 | annual premium $240
    - monthly premium $25 | annual premium $300
    - monthly premium $30 | annual premium $360
    - monthly premium $40 | annual premium $480
    - monthly premium $50 | annual premium $600
    - monthly premium $75 | annual premium $900
    - monthly premium $100 | annual premium $1,200
    - monthly premium $125 | annual premium $1,500
    - monthly premium $150 | annual premium $1,800
    - monthly premium $200 | annual premium $2,400
    - monthly premium $300 | annual premium $3,600
    - monthly premium $400 | annual premium $4,800
    - monthly premium $500 | annual premium $6,000
30. If you own your home, do you currently have a flood insurance policy covering your home? Keep in mind that standard homeowner's insurance policies do not include flood insurance and it must be purchased separately.
    - Yes
    - No
    - Unsure
    - I am not a homeowner/not applicable
31. Suppose your city or town is struck by a major hurricane or flood in 2018 that causes widespread devastation. If that happens, would you think another hurricane or flood is more likely or less likely to come in 2019?
    - More likely
    - About the same
    - Less likely

Memory of past floods

32. What is the elevation of your home in feet? You can check by entering your exact street address here: https://www.whatisoutelevation.com/
33. Have you experienced a flood at your current residence where water entered your home and/or damaged your property?
34. What year did your home flood? Enter only the most recent year if you have experienced multiple floods.
35. Try to remember the last time the town or region where you live experienced a major hurricane or major flood. This may have occurred before you moved to the area or even before you were born.
    - I can remember, or have heard about, the last major hurricane or flood in my area
    - I don't know of any major hurricanes or floods that have ever struck my area
36. If applicable, what was the name of the hurricane or flood that hit your region?
37. Enter the approximate year of the most recent hurricane or flood that hit your area.

Attitudes about climate change

38. In your view, the risk of your home flooding twenty years in the future will be:
    - Much greater than today
    - Somewhat greater
    - About the same
    - somewhat less
    - Much less than today
39. In your own words, why did you answer the above question the way you did?
40. At this point we'd also like to ask you about your opinions on climate change. Do you believe that the global climate is changing?
    - Yes, it is changing
    - Unsure
    - No, it is not changing
41. To the best of your knowledge, how much have global temperatures increased in the last 200 years? Enter the number of degrees Fahrenheit.
42. What are your reasons for completing this survey?

A7 Appendix figures and tables

Figure A1 · FEMA flood map for Miami, Florida

A screen capture of the online flood map utility available from FEMA, featuring a map of downtown Miami, Florida. The blue area is the 100-year flood zone, and the orange area (visible in a small strip in the top-center of the image) is the 500-year flood zone. Note that the blue areas is further divided into VE, AE, and AH zones. Highly detailed, localized flood maps are available at https://msc.fema.gov/portal/search.

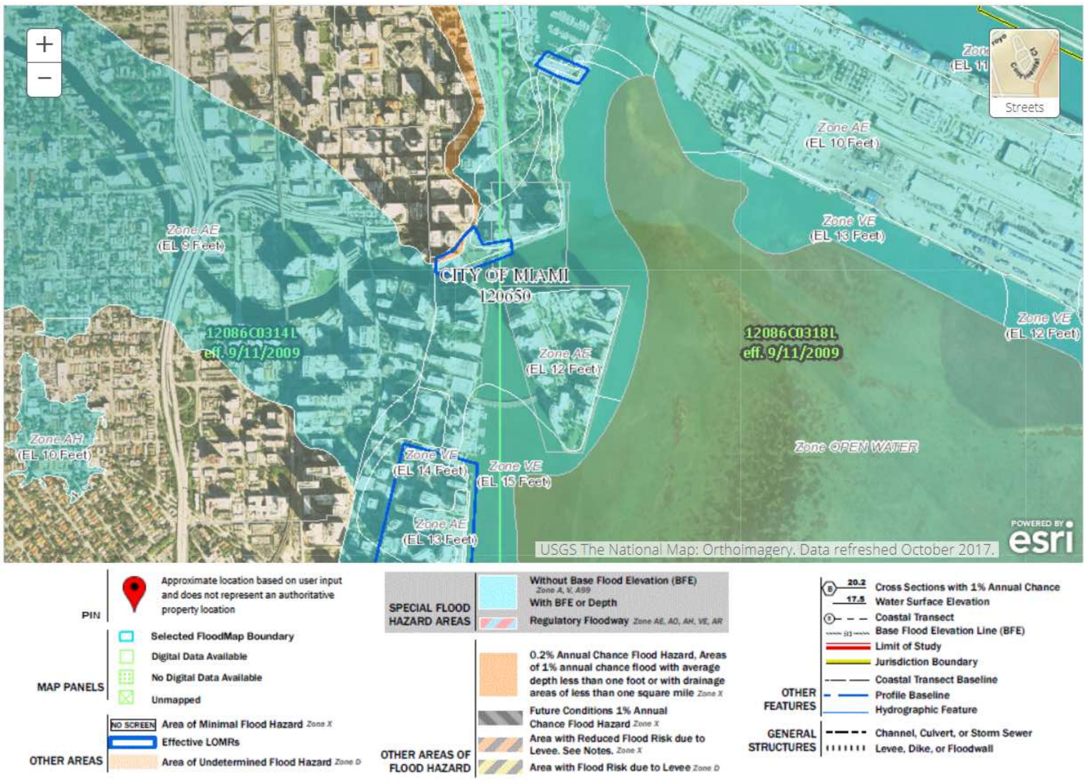


Table A1 · Robustness of hedonic models

	(1)	(2)	(3)	(4)
		Disaster pressure threshold		
	Spec from Table 1.8	1 year	2 years	3 years
Property characteristics ($\beta_{x,1}$)				
Square footage (log)	0.874***	0.895***	0.877***	0.885***
	(0.0120)	(0.00900)	(0.0115)	(0.0150)
Year built/renovated	0.00396***	0.00244***	0.00294***	0.00449***
	(0.000391)	(0.000314)	(0.000396)	(0.000496)
Pool indicator	0.129***	0.168***	0.136***	0.102***
	(0.00929)	(0.00730)	(0.00938)	(0.0118)
Garage indicator	0.123***	0.191***	0.138***	0.0657***
	(0.0116)	(0.0117)	(0.0111)	(0.0153)
Hydrology main effects ($\beta_{w,1}$)				
Frontage	0.185***	0.168***	0.175***	0.184***
	(0.0320)	(0.0302)	(0.0317)	(0.0363)
< 6-ft SLR zone	0.00550	-0.0353**	-0.0127	0.00489
	(0.0186)	(0.0170)	(0.0191)	(0.0230)
6-ft SLR zone	0.0456***	0.0141	0.0455***	0.0734***
	(0.0135)	(0.0149)	(0.0149)	(0.0197)
Hydrology interaction effects ($\beta_{w,2}$)				
Frontage	0.000656	0.000160	-0.000511	-0.000982
	(0.00161)	(0.00137)	(0.00163)	(0.00184)
< 6-ft SLR zone	-0.00445***	-0.00216**	-0.00296**	-0.00304**
	(0.00142)	(0.00108)	(0.00130)	(0.00136)
6-ft SLR zone	-0.00542***	-0.00465***	-0.00520***	-0.00533***
	(0.00142)	(0.00126)	(0.00155)	(0.00165)
Number of census tract FE	3,068	3,065	3,065	3,065
Quarter-of-sample FE	✓	✓	✓	✓
Observations	2,342,896	1,878,851	1,878,851	1,878,851
R-squared	0.406	0.405	0.406	0.408

OLS estimates with robust standard errors clustered at the census tract in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is the log of sales price in nominal dollars. Frontage properties are those within 50 m of the water. The results in Column 1 are from the same model as Column 5 in Table 1.8. Disaster pressure in Column 1 is the percentage of buyers in an MSA-month whose home counties experienced a flood event in the 24 months leading up to the transaction. Disaster pressure is defined with various thresholds in Columns 2, 3, and 4. Column 1 excludes Florida markets with a disaster in the past 24 months. Columns 2, 3, and 4 exclude Florida markets with a disaster in the past 36 months. This accounts for the difference in sample size and census tract count, and the minor discrepancies between Columns 1 and 3.

Figure A2 · Illustration of treatment definition

The figure illustrates how regions are assigned to treatment based on FEMA disaster or emergency declarations and elevated flood claims. Consider a hypothetical flooding disaster on the Monongahela River which affects and leads to a disaster declaration for Allegheny County, Pennsylvania (highlighted in light blue). ZIP codes with an abnormal number of flood claims reported in the month of the storm are highlighted in dark blue; they would not be included in some analysis, such as Column 2 in Table 1.5. Only the ZIP codes outside the damage zone, but in the county with the reported flood disaster, are considered treated. Adjacent counties that were not subject to a disaster declaration are not considered treated.

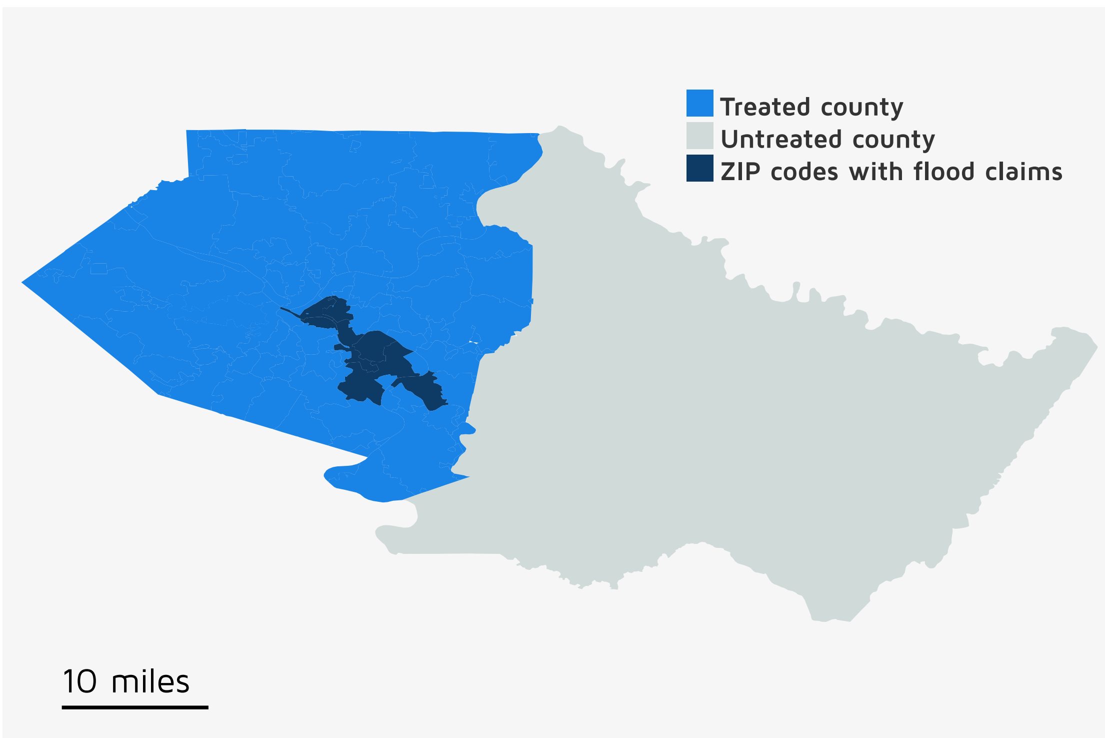


Figure A3 · QQ plots for Tampa MSA

The following are QQ plots produced by the MatchIt package in R (Stuart et al. 2011) for the Tampa MSA sample. Figure 1.10 shows a comparison of standardized means for this same sample. The QQ plots show that the genetic matching algorithm also matches higher moments very closely.

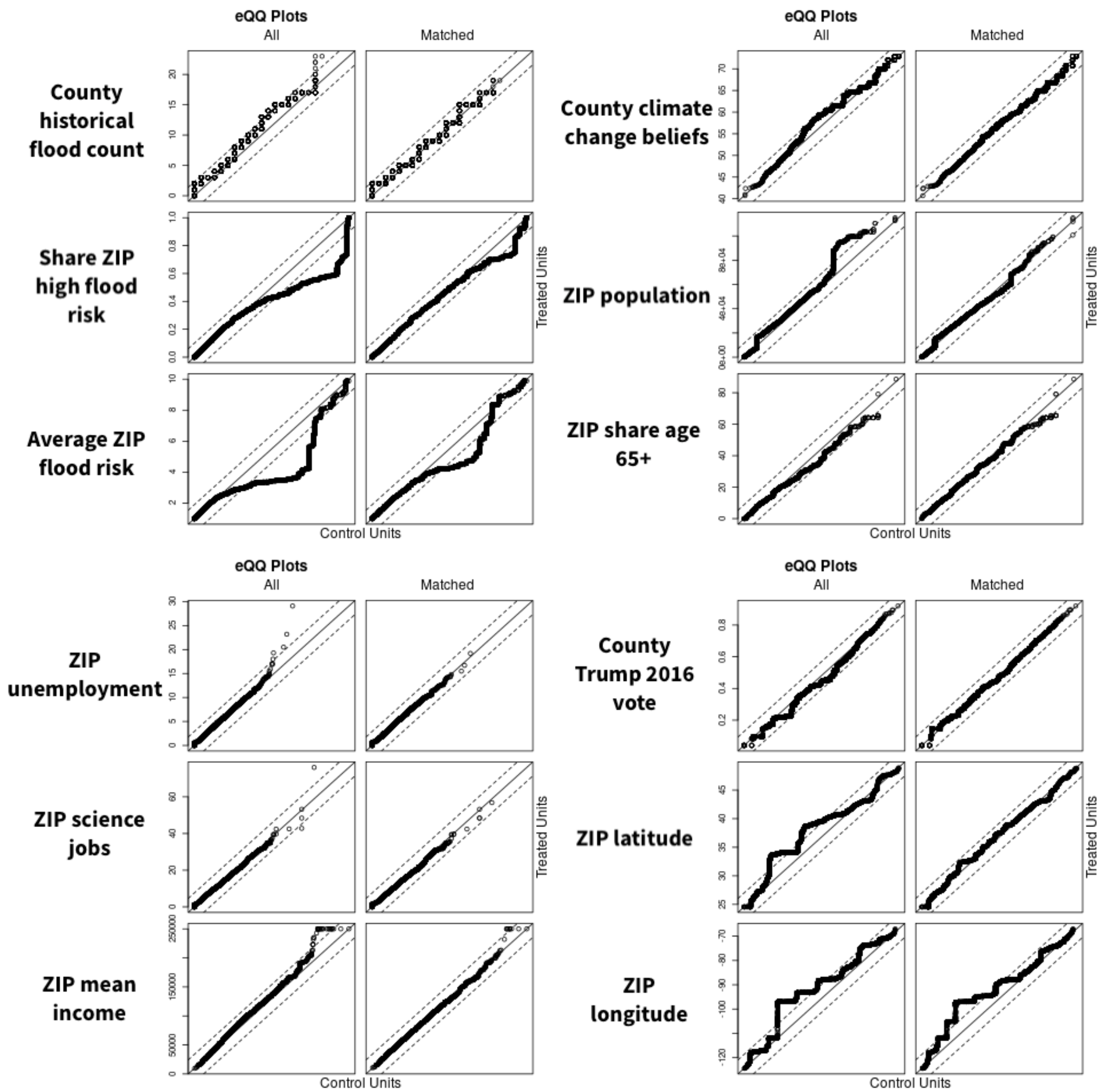


Figure A4 · Principal component analysis of Florida MSA catchment areas

A scatterplot where each MSA-year is plotted according to its first two components in a PCA analysis. The input vectors were the share of sales in each MSA-year for each of the 49 states outside Florida. Based on the loadings of individual states within each component, the first component seems to represent a south-north continuum from left to right in the figure. The second component represents a continuum from Midwestern states in the top of the figure to coastal states in the bottom. The close clustering within MSA across years suggests that the regional flavor of each MSA's arrivals is very persistent over time. The three MSAs located in the Florida Panhandle (Destin, Pensacola, Panama City) are quite isolated from the other MSAs due to distinctive and heavy inflows from southern states close to Florida. The state's largest cities are clustered at the bottom of the figure, while the retirement communities on the Gulf Coast (Ft. Myers, Naples, Punta Gorda, Sarasota) are clustered at the top. This reflects the relatively large inflows from Midwestern states to the Gulf Coast region.

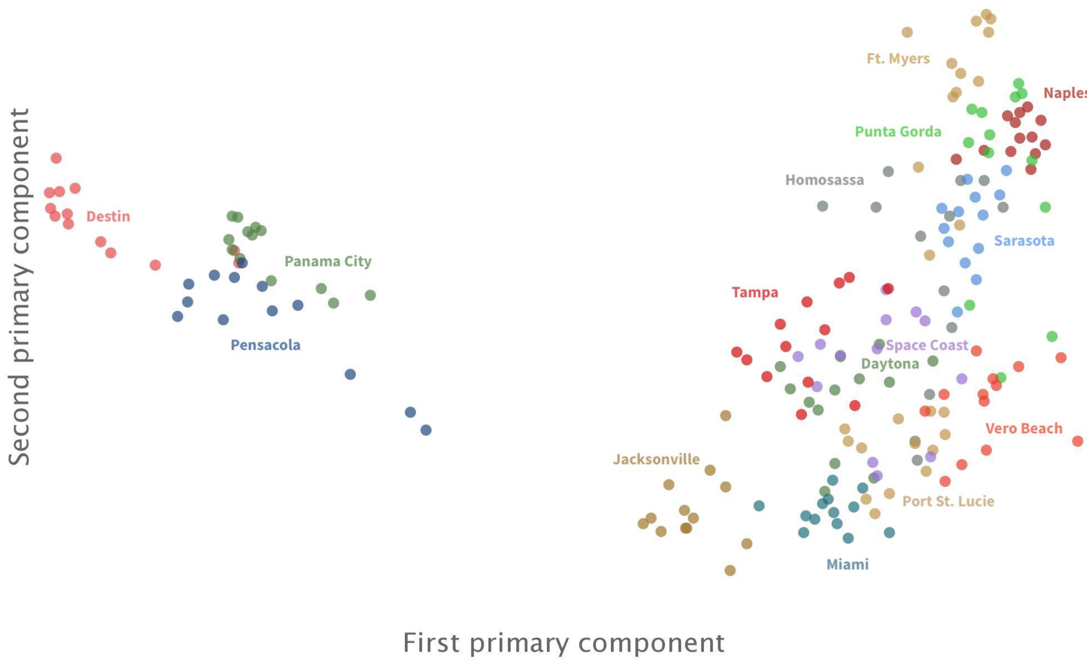


Figure A5 · Example addition task from lab paradigm

Example grid for addition task used in lab paradigm. Participants were given the following explanation: "Every grid you will see contains several numbers ranging between 0.00 to 10.00. To solve a grid, you must find the unique pair of numbers within the grid that sum to exactly 10.00."

8.45	2.42	8.98
3.38	9.75	8.87
9.43	4.41	1.02

Table A2 · Placebo test of flood map exposure

Variable	Control group	Experimental group	p-value (H_0: equal means)
Past events			
Carries flood insurance	0.50	0.46	0.30
	(0.02)	(0.04)	
Past flood in home	0.18	0.15	0.27
	(0.01)	(0.02)	
Attitudinal questions			
Perceived cost of flood ($ thousands)	62.45	59.91	0.79
	(8.04)	(5.46)	
Government flood support	33.13%	32.02%	0.49
	(0.79)	(1.39)	
Any memory of past local hurricane	0.81	0.76	0.05
	(0.01)	(0.02)	
Believe climate change is occurring	0.80	0.79	0.78
	(0.01)	(0.02)	
Anchoring test			
Listed flood likelihood of 1% or 0.2%	0.038	0.036	0.84
	(0.01)	(0.01)	

Sample mean values with standard errors in parentheses. Government flood support is the fraction of flood damages that respondents believe would be covered by the government (See Appendix B, Question 28). A two-sample t-test of the equality of means with unequal variances was performed on each variable, and the p-value for each test is listed in the right-most column. See Section 3.3.3 for an explanation of these tests.

Table A3 · Effect of flood map exposure with municipality-level flood data

	Perceived 10-year flood likelihood (log)			Worry about floods (1-10 scale)			WTP for flood insurance		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Map exposure	-0.524***	-0.503***	-0.652***	-0.674***	-0.543***	-0.745***	-134.3	-38.62	36.84
	(0.146)	(0.154)	(0.208)	(0.130)	(0.141)	(0.212)	(86.04)	(87.85)	(132.5)
Home elevation (ft)		-0.005***	-0.005		-0.007***	-0.0116***		-1.608	-3.210
		(0.001)	(0.003)		(0.001)	(0.003)		(1.512)	(2.183)
Lives walking dist. to water		0.716***	0.732***		0.984***	0.899***		67.18	-1.859
		(0.115)	(0.162)		(0.105)	(0.135)		(61.78)	(95.11)
Municipality floods since 1990		-0.0157	1.063***		0.145**	1.848***		10.64	-48.38
		(0.0439)	(0.264)		(0.0556)	(0.279)		(19.01)	(152.4)
Past flood in home		1.505***	1.496***		2.005***	1.709***		541.8***	562.7***
		(0.117)	(0.132)		(0.142)	(0.213)		(116.1)	(174.6)
Climate believer		0.391***	0.361*		0.617***	0.602***		54.39	182.3*
		(0.116)	(0.183)		(0.141)	(0.202)		(84.19)	(99.74)
Hydrological controls		✓	✓		✓	✓		✓	✓
Demographic controls		✓	✓		✓	✓		✓	✓
Municipality fixed effects			✓			✓			✓
Observations	1,582	1,300	1,300	1,582	1,300	1,300	1,429	1,180	1,180
Adjusted R^2	0.00952	0.166	0.200	0.0122	0.212	0.228	0.00159	0.0509	0.127

OLS estimates with standard errors clustered at the county level in parentheses, *** p<0.01, ** p<0.05, * p<0.1. These specifications are analogous to those shown in Table 3.5, but with municipality-level flood history data dating back to 1990. Some respondents could not be matched to municipalities in this dataset so the sample size is somewhat restricted. Hydrological controls (shown) comprise home elevation, proximity to the water, municipality-level historical flooding data, information on past floods in the home, and climate change beliefs. Demographic controls (not shown) comprise gender, homeowner status, tenure in current home, and categorical bins for age, race, income, educational attainment, home type, and risk preference. Specifications 3, 6, and 9 include municipality fixed effects. A small number of observations with missing or unreliable data for demographic controls and are omitted from some regressions; see Table 3.3.

Table A4 · Robustness check: Effect of map exposure with attention screen

	Perceived 10-year flood likelihood (log)			Worry about floods (1-10 scale)			WTP for flood insurance		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Map exposure	-0.518***	-0.487***	-0.648**	-0.632***	-0.568***	-0.740***	-77.95	-76.80	-19.05
	(0.175)	(0.178)	(0.304)	(0.149)	(0.142)	(0.240)	(70.83)	(65.93)	(118.9)
Home elevation (ft)		-0.00625***	-0.00331		-0.0115***	-0.00891**		-2.557**	-0.571
		(0.00191)	(0.00433)		(0.00190)	(0.00396)		(1.105)	(2.853)
Live walking dist. to water		0.735***	0.812***		0.998***	0.917***		92.45*	26.32
		(0.115)	(0.225)		(0.112)	(0.173)		(53.47)	(98.80)
County floods since 1965		0.00698	0.721***		0.0631	2.351***		11.40	60.40
		(0.0239)	(0.270)		(0.0412)	(0.198)		(11.58)	(134.8)
Past flood in home		1.711***	1.502***		1.968***	1.624***		311.2***	218.0
		(0.170)	(0.251)		(0.211)	(0.332)		(101.6)	(150.2)
Climate believer		0.449***	0.503**		0.717***	0.855***		-63.64	-6.154
		(0.136)	(0.239)		(0.164)	(0.282)		(90.78)	(134.1)
Hydrological controls		✓	✓		✓	✓		✓	✓
Demographic controls		✓	✓		✓	✓		✓	✓
Municipality fixed effects			✓			✓			✓
Observations	1,203	1,191	1,191	1,203	1,191	1,191	1,142	1,132	1,132
Adjusted R^2	0.00850	0.152	0.141	0.0102	0.185	0.212	2.11e-05	0.0246	0.0940

OLS estimates with standard errors clustered at the county level in parentheses, *** p<0.01, ** p<0.05, * p<0.1. These specifications are analogous to those shown in Table 3.5, but with respondents who failed attention checks excluded from the sample. Of the full sample, 22% failed the general attention question and 12% of the experimental group failed an experiment-specific attention question. Hydrological controls (shown) comprise home elevation, proximity to the water, municipality-level historical flooding data, information on past floods in the home, and climate change beliefs. Demographic controls (not shown) comprise gender, homeowner status, tenure in current home, and categorical bins for age, race, income, educational attainment, home type, and risk preference. Specifications 3, 6, and 9 include municipality fixed effects. A small number of observations with missing or unreliable data for demographic controls and are omitted from some regressions; see Table 3.3.

Table A5 · Robustness check: Map interaction effect with attention screen

	Perceived flood likelihood		WTP for flood insurance	
	(1)	(2)	(3)	(4)
Main effects				
Home elevation (ft)	-0.00353*	-0.00251	-2.125*	-1.797
	(0.00202)	(0.00178)	(1.196)	(1.492)
Lives walking dist. to water	0.775***	0.668***	85.90	101.3
	(0.132)	(0.149)	(67.59)	(72.78)
Municipality floods since 1990		-0.00613		20.19
		(0.0464)		(22.94)
County floods since 1965	-0.0192	-0.0263	11.87	13.54
	(0.0220)	(0.0235)	(13.69)	(15.25)
Past flood in home	1.887***	1.793***	390.3***	414.7***
	(0.188)	(0.202)	(123.7)	(136.6)
Climate believer	0.553***	0.539***	-68.24	-38.85
	(0.154)	(0.171)	(93.86)	(102.0)
Interaction effects				
Home elevation * Map	-0.0165***	-0.0157***	-2.640	-1.325
	(0.00468)	(0.00521)	(2.268)	(2.734)
Lives walking distance * Map	-0.232	-0.120	2.267	-3.500
	(0.324)	(0.382)	(149.3)	(167.9)
Municipality floods * Map		0.0148		-14.36
		(0.119)		(45.53)
County floods * Map	0.133**	0.130*	2.737	-18.17
	(0.0530)	(0.0679)	(29.80)	(24.63)
Past flood in home * Map	-0.868**	-0.808*	-395.7**	-391.0**
	(0.437)	(0.483)	(173.0)	(156.6)
Climate believer * Map	believer	-0.556	-5.396	124.2
	(0.330)	(0.364)	(142.7)	(148.7)
Demographic controls	✓	✓	✓	✓
Municipality-level data		✓		✓
Observations	1,191	998	1,132	945
Adjusted R^2	0.162	0.144	0.0253	0.0184

OLS estimates with standard errors clustered at the county level in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$. These specifications are analogous to those shown in Table 3.5, but with respondents who failed attention checks excluded from the sample. 22% of the sample failed the general attention question and 12% of the experimental group failed an experiment-specific attention question. Perceived flood likelihood is the natural log of the respondent's stated belief that her home will flood in the next ten years. Willingness to pay for flood insurance is expressed in terms of annual premia for a $100,000 policy. Demographic controls (not shown) comprise gender, homeowner status, tenure in current home, and categorical bins for age, race, income, educational attainment, home type, and risk preference. Some specifications include municipality-level flood history from FEMA, which is not available for some respondents because their municipalities do not appear in the FEMA database. Municipality fixed effects are not included in these specifications; see text for further discussion.

Table A6 · Robustness check: Effect of map exposure without imputed elevations

	Perceived 10-year flood likelihood (log)			Worry about floods (1-10 scale)			WTP for flood insurance		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Map exposure	-0.567***	-0.586***	-0.751***	-0.615***	-0.576***	-0.801***	-141.6*	-123.2	-49.57
	(0.175)	(0.176)	(0.278)	(0.159)	(0.146)	(0.241)	(84.25)	(78.89)	(135.7)
Home elevation (ft)		-0.00640***	-0.00389		-0.0101***	-0.0106**		-2.521**	-3.676
		(0.00194)	(0.00480)		(0.00172)	(0.00415)		(1.028)	(2.385)
Lives walking dist. to water		0.833***	0.834***		1.108***	1.030***		82.85	53.26
		(0.120)	(0.213)		(0.137)	(0.218)		(61.86)	(95.37)
County floods since 1965		0.00738	0.671***		0.0660	1.648***		28.51**	-19.15
		(0.0276)	(0.250)		(0.0417)	(0.238)		(11.78)	(123.7)
Past flood in home		1.631***	1.444***		2.096***	1.628***		380.5***	281.1
		(0.131)	(0.183)		(0.193)	(0.326)		(112.9)	(173.8)
Climate believer		0.311**	0.380		0.602***	0.761***		-45.91	60.93
		(0.135)	(0.239)		(0.157)	(0.229)		(93.48)	(141.4)
Hydrological controls		✓	✓		✓	✓		✓	✓
Demographic controls		✓	✓		✓	✓		✓	✓
Municipality fixed effects			✓			✓			✓
Observations	1,197	1,185	1,185	1,197	1,185	1,185	1,097	1,088	1,088
Adjusted R^2	0.0111	0.190	0.211	0.0101	0.230	0.259	0.00194	0.0421	0.130

OLS estimates with standard errors clustered at the county level in parentheses, *** p<0.01, ** p<0.05, * p<0.1. These specifications are analogous to those shown in Table 3.5, but with respondents who provided implausible home elevation data excluded (rather than having their elevations imputed from their ZIP codes). Hydrological controls (shown) comprise home elevation, proximity to the water, municipality-level historical flooding data, information on past floods in the home, and climate change beliefs. Demographic controls (not shown) comprise gender, homeowner status, tenure in current home, and categorical bins for age, race, income, educational attainment, home type, and risk preference. Specifications 3, 6, and 9 include municipality fixed effects. A small number of observations with missing or unreliable data for demographic controls and are omitted from some regressions; see Table 3.3.

Figure A6 · Relationship between disaster pressure and same-address listing

OLS estimates of β from the specifications in Appendix A2, with robust 95% confidence intervals.

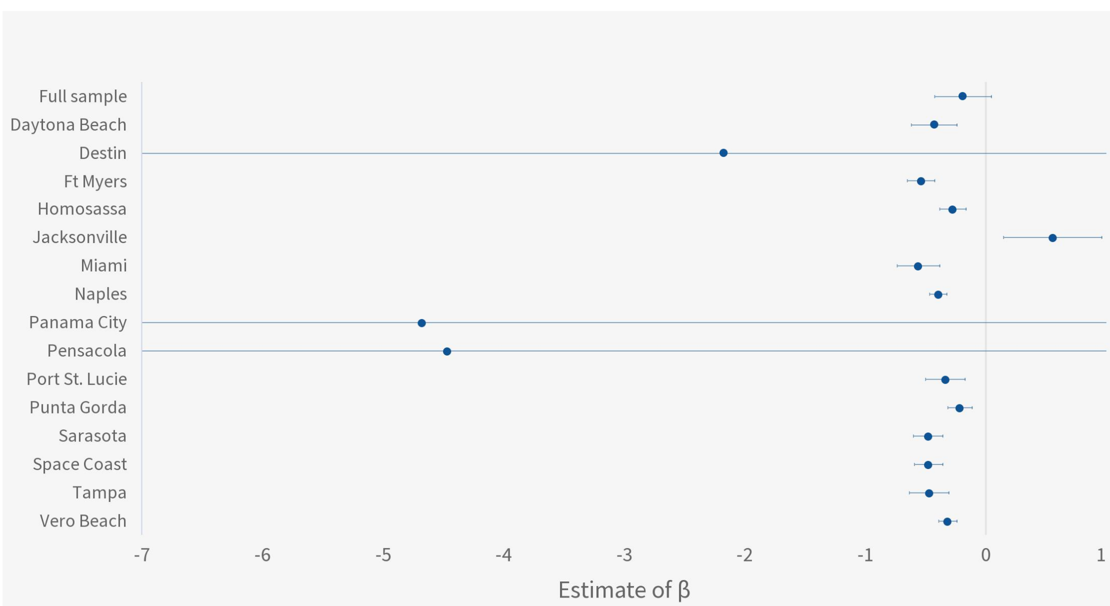


Table A7 · Robustness check: Map interaction effect without imputed elevations

	Perceived flood likelihood		WTP for flood insurance	
	(1)	(2)	(3)	(4)
Main effects				
Home elevation (ft)	-0.00399*	-0.00347*	-2.362**	-1.999
	(0.00213)	(0.00182)	(1.095)	(1.439)
Lives walking dist. to water	0.811***	0.741***	82.18	109.4
	(0.143)	(0.158)	(78.45)	(78.39)
Municipality floods since 1990		-0.0205		-0.153
		(0.0498)		(24.09)
County floods since 1965	-0.00993	-0.00391	37.37***	33.31**
	(0.0239)	(0.0265)	(13.87)	(15.37)
Past flood in home	1.725***	1.639***	412.5***	474.2***
	(0.148)	(0.163)	(141.8)	(154.4)
Climate believer	0.456***	0.526***	-83.02	-51.57
	(0.160)	(0.169)	(97.87)	(98.96)
Interaction effects				
Home elevation * Map	-0.0140***	-0.0152***	-1.257	-0.802
	(0.00443)	(0.00535)	(2.004)	(2.763)
Lives walking distance * Map	-0.00515	-0.00243	-33.62	-69.61
	(0.265)	(0.310)	(130.1)	(140.4)
Municipality floods * Map		0.0639		54.54
		(0.131)		(44.95)
County floods * Map	0.0801	0.0599	-36.63	-67.27***
	(0.0614)	(0.0761)	(27.39)	(21.78)
Past flood in home * Map	-0.490	-0.419	-115.8	-53.19
	(0.344)	(0.374)	(287.8)	(298.8)
Climate believer * Map	-0.585*	-0.690**	128.6	187.5
	(0.297)	(0.308)	(127.0)	(138.3)
Demographic controls	✓	✓	✓	✓
Municipality-level data		✓		✓
Observations	1,185	997	1,088	914
Adjusted R^2	0.198	0.189	0.0412	0.0459

OLS estimates with standard errors clustered at the county level in parentheses, *** p<0.01, ** p<0.05, * p<0.1. These specifications are analogous to those shown in Table 3.6, but with respondents who provided implausible home elevation data excluded (rather than having their elevations imputed from their ZIP codes). Perceived flood likelihood is the natural log of the respondent's stated belief that her home will flood in the next ten years. Willingness to pay for flood insurance is expressed in terms of annual premia for a $100,000 policy. Demographic controls (not shown) comprise gender, homeowner status, tenure in current home, and categorical bins for age, race, income, educational attainment, home type, and risk preference. Some specifications include municipality-level flood history from FEMA, which is not available for some respondents because their municipalities do not appear in the FEMA database. Municipality fixed effects are not included in these specifications; see text for further discussion.