Essays on the Economics of Environmental and Health Risk

by

Abigail J.H. Ostriker

B.S., Massachusetts Institute of Technology (2016)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2023

Authored by: Abigail J.H. Ostriker
Department of Economics
May 12, 2023

Certified by: Amy Finkelstein
John and Jennie S. MacDonald Professor of Economics
Thesis Supervisor

Certified by: Clare Balboni
Assistant Professor of Economics
Thesis Supervisor

Accepted by: Abhijit Banerjee
Ford International Professor of Economics
Chairman, Departmentmental Committee on Graduate Studies

Essays on the Economics of Environmental and Health Risk

by

Abigail J.H. Ostriker

Submitted to the Department of Economics
on May 12, 2023, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Individuals face many types of risk, including to their property, finances, and health. In this thesis, I study a variety of strategies for mitigating those risks, in the context of health shocks and natural disasters. In the health context, risk of disease onset is mitigated in some cases with screening programs, with governments recommending an age to begin screening. In the environmental context, risk of physical destruction is mitigated with property insurance, which can be purchased from the federal government (for floods) or private insurers (for wildfires). In the first two essays, I study the effectiveness of government policies recommending screening ages and regulating floodplain development via rules embedded in the National Flood Insurance Program. In the third essay, I investigate how higher property insurance prices affect housing markets. All three essays highlight the importance of government policies and insurance products for mitigating personal risk.

JEL Codes: I18, Q54, R31

Thesis Supervisor: Amy Finkelstein
Title: John and Jennie S. MacDonald Professor of Economics

Thesis Supervisor: Clare Balboni
Title: Assistant Professor of Economics

Acknowledgments

This thesis would not have been possible without the invaluable support of countless advisors, friends, and family members, as well as the broader MIT Economics community.

I am deeply indebted to my official graduate advisors, Amy Finkelstein, Clare Balboni, and Ben Olken, as well as my first-ever mentor in economics, Nikhil Agarwal. Nikhil introduced me to economics research with patience and attention to my interests, and his nudge to consider graduate school (and good advice about how to do so) changed the course of my career. Clare nurtured my budding interest in environmental economics, provided detailed feedback on innumerable drafts, and gave me what feels like dozens of pep talks through the low points of a third year made even more challenging by Covid. Ben challenged my research frameworks, helped me communicate the forest instead of the trees, and did it all with a level of constructive argumentativeness I would never have thought humanly possible. Finally, it's impossible to overstate the influence Amy has exerted on me, from imprinting me with her preferred project workflow to giving me (or saddling me with) her taste in research. Any time I ask, "But what is the market friction?" or, "Why don't these two N's match?", that's Amy speaking. Beyond her sage advice and terrible puns, I am forever grateful to Amy for her persistent kindness and unstinting generosity in matters large and small.

My other coauthors have also been my teachers, and I am grateful for their drive, thoughtfulness, and good cheer. Thank you to Liran Einav for walking me through my first structural model estimation, to Heidi Williams for sharing her medical expertise, and to Tamar Oostrom for being a light in the wilderness of Medicare data. I am especially grateful to Anna Russo for five years of exceptional coauthorship, in which she countered my native pessimism with her own enthusiasm, held us to the highest standards, and exhibited unfailing creativity and tenacity, on top of being a great friend.

My other great friends also deserve my gratitude. I could always rely on Maya Bidanda, Lucy Page, and Hannah Ruebeck, individually and collectively, for a listening ear, a quick proofread, or a laugh. Thank you for helping smooth a million tiny bumps in the road, and also some bigger ones. I also want to thank Annie Dunn for years of friendship and practical advice, and for thinking my work in economics is the least interesting thing about me.

The isolation of the Covid years has given me a profound appreciation for serendipitous in-person interactions, however brief, so I want to thank every individual in the MIT Economics department who has ever simply been willing to say hi. In particular, I am grateful to the regular participants at the MIT public finance lunches and the MIT and Harvard environmental student groups for enthusiastically countenancing early-stage ideas, troubleshooting tricky issues, and forming a community I loved being part of.

There are some practicalities without which no research would ever get done. I am deeply grateful to Andy Dorner for solving every computer problem I could throw at him, even the weird errors he had never seen before, with a response time that verged on instantaneous. I am also grateful to Jerry Hausman, Carl Shapiro, George and Obie Shultz, and the National Science Foundation for financial support.

Lastly, I owe a great debt to the many members of my family who paved the way for this achievement and gave me their limitless love and support. Each of my grandparents – Karl and Sue Bottigheimer, and Jerry and Alicia Ostriker – devoted their careers to expanding humanity's understanding of the world and our place in it. Thank you for teaching me the value of education and curiosity, and helping me to follow in your footsteps. My parents, Nat and Eve, and my sister, Naomi, shaped my interests and priorities growing up and to this day. I am grateful to my father for inspiring me to study public affairs, and my mother for giving me the mathematical foundation I needed for this work. Finally, to my husband Bart, who saw me through it all: thank you for grounding me, hearing me, and believing in me, during the Ph.D. and always.

Contents

List of Figures

List of Tables

Chapter 1

Screening and Selection: The Case of Mammograms

By Liran Einav, Amy Finkelstein, Tamar Oostrom, Abigail Ostriker, and Heidi Williams*

Abstract. We analyze selection into screening in the context of recommendations that breast cancer screening start at age 40. Combining medical claims with a clinical oncology model, we document that compliers with the recommendation are less likely to have cancer than younger women who select into screening or women who never screen. We show this selection is quantitatively important: shifting the recommendation from age 40 to 45 results in three times as many deaths if compliers were randomly selected than under the estimated patterns of selection. The results highlight the importance of considering characteristics of compliers when making and designing recommendations.(*JEL* I11, I18)

*Einav: Department of Economics, Stanford University, and the National Bureau of Economic Research, leinav@stanford.edu; Finkelstein: Department of Economics, Massachusetts Institute of Technology, and the National Bureau of Economic Research, afink@mit.edu; Oostrom: Department of Economics, Ohio State University, oostrom.1@osu.edu; Ostriker: Department of Economics, Massachusetts Institute of Technology, ostriker@mit.edu; Williams: Department of Economics, Stanford University, and the National Bureau of Economic Research, hlwill@stanford.edu.

We are grateful to Leila Agha, Emily Oster, Doug Owens, three anonymous referees, and participants in the Dartmouth/NIA P01 Research Meeting and the NBER Health Care Summer Institute for helpful comments, and to the Laura and John Arnold Foundation for financial support. This material is based upon work supported by the National Institute on Aging through Grant Number T32-AG000186 and the National Science Foundation Graduate Fellowship Program under Grant Number 1122374 (Oostrom). The authors acknowledge the assistance of the Health Care Cost Institute (HCCI) and its data contributors, Aetna, Humana, and UnitedHealthcare, in providing the claims data analyzed in this study. This study also used the linked SEER-Medicare database. The interpretation and reporting of these data are the sole responsibility of the authors. The authors acknowledge the efforts of the National Cancer Institute; the Office of Research, Development and Information, CMS; Information Management Services (IMS), Inc.; and the Surveillance, Epidemiology, and End Results (SEER) Program tumor registries in the creation of the SEER-Medicare database.

Research in this chapter previously appeared in Einav et al. (2020). Please see the online appendix of that paper for all references to the appendix.

Whether and when to recommend screening for potential diseases is a highly controversial and evolving policy area, with active academic research.[1] Much of the debate—both in public policy and in academia—centers on the causal impact of screening for a typical individual covered by the recommendation. Estimating this causal impact is challenging for several well-known reasons. First, there are the usual challenges to causal inference. Second, many of the potential costs and benefits of screening are difficult to measure and to monetize.[2] In this paper, we highlight another important—and, we believe, overlooked—challenge in analyzing and designing screening recommendations: the typical individual covered by a recommendation may be very different from the typical individual who responds to the recommendation. As a result, the estimated impact of screening for a randomly-selected individual may be quite different from the impact for an affected individual.

We explore this distinction in the context of the current controversy over whether to recommend annual mammograms for women starting at age 40. Results from randomized trials have consistently failed to show statistically significant mortality benefits of mammograms for women in their 40s. In 2009, these results prompted the US Preventive Services Task Force (USPTF) to change its recommendation for routine mammograms to begin at age 50 rather than at age 40. This change generated substantial public controversy (Kolata 2009; Saad 2009; Berry 2013).

This debate has focused on the costs and benefits of mammograms for typical ("average-risk") 40-year-old women, with little attention paid to what types of women respond to a screening recommendation and whether the costs and benefits for them may differ from the average woman. To investigate the type of women who respond, we draw on two primary data sources. The first is insurance claims data on mammogram choices and their results (negative, false positive, or true positive) for privately-insured women aged 35-50 from the Health Care Cost Institute (HCCI). The second is cancer registry data, from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) database, on the size and stage of detected tumors for women aged 35-50 who were diagnosed with breast cancer. We supplement some of the descriptive analyses with additional information from the Behavioral Risk Factor Surveillance System Survey (BRFSS), which allows us to observe additional health behaviors and demographics of women who do and do not receive mammograms at various ages.

The visual evidence shows sharp and pronounced changes in behavior and outcomes at age 40. There is a 25-percentage-point jump in the annual mammogram rate at age 40, from 10 percent to 35 percent of women. We then compare characteristics of the women who respond to the recommendation for a mammogram (i.e. "compliers" in the terminology of Angrist, Imbens, and Rubin 1996) to characteristics of always-takers (i.e. women who choose mammograms even in the absence of the recommendation, which is before age 40). We find that compliers have a lower incidence of cancer than always-takers: there is a roughly 30 percent decline (from 0.84% to 0.56%) in the share of screened women diagnosed with cancer (i.e. true positives) at age 40. Given the high rate of false positives—about 90 percent of initial positive mammograms turn out to be false positives—the sharp increase in the mammogram rate at age 40 translates into a substantial increase in the number of women experiencing false positives, from about 10 per thousand women to about 40 per thousand women. This is consistent with false positives being a key concern that motivated moving the recommended age of beginning mammography from 40 to 50 (Nelson et al. 2009). Moreover, among those

[1]For example, Welch, Schwartz, and Woloshin (2011) argue that although many medical conditions—such as high blood pressure, elevated blood glucose levels, low bone density, and high cholesterol—benefit from treatment, there has been a trend over time towards widespread use of medical screening tests and increasingly low diagnostic thresholds that recommend treating patients for whom the benefits from treatments are quite small. By contrast, Maciosek et al. (2010) review these same screening efforts and conclude that they save a large number of lives at relatively low cost.

[2]The costs and benefits of screening include monetary costs, clinical outcomes, discomfort from unnecessary procedures, and psychological effects induced by the screening process, including pre-screening apprehension and anxiety due to false positives (e.g., Brett et al. 2005; Nelson et al. 2009; Welch and Passow 2014; Ong and Mandl 2015; Welch 2015).

diagnosed with cancer, the registry data show a sharp decline in the average tumor's stage and size starting at age 40, compared to earlier ages. For example, the share of detected tumors that are in a late stage (invasive tumors) as opposed to early stage (in-situ tumors) falls by about 6 percentage points (or 7 percent) at age 40.

These descriptive results indicate that women who respond to the recommendation for a mammogram have lower risk of cancer than those who seek mammograms in the absence of the recommendation. For non-cancer characteristics, we can also compare compliers to never-takers (women who do not get mammograms even once the recommendation is in effect). We find that, relative to never-takers, compliers are more likely to undertake other types of recommended preventive care, such as cervical cancer screening tests and flu shots. This pattern is consistent with findings that when a health behavior is recommended, those who comply with the recommendation tend to exhibit other positive health behaviors (Oster 2020). It also echoes the observation that women who comply with assignment to mammograms in an RCT setting are healthier than never-takers (Kowalski 2019).[3]

To assess the implications of these findings and to quantify costs and health outcomes under various counterfactual selection scenarios, we specify a model of mammogram demand that is a function of a woman's age, her (undiagnosed) cancer type (no cancer, in-situ, or invasive), and whether or not a mammogram is recommended at her age. We estimate the model by method of moments, using two key inputs. First, we leverage our data on the observed patterns of mammogram decisions and mammogram outcomes (specifically, cancer type) for women by age. Second, we bring in a clinical oncology model of the underlying rate of onset of breast cancer by age, as well as the cancer's clinical progression in the absence of detection and treatment.

The clinical model allows us to estimate the cancer characteristics of never-takers. In the absence of a clinical model, these cancer characteristics are inherently difficult (or impossible) to observe: cancer incidence is not observed in the non-screened population, and almost all detected cancer is treated immediately upon detection. The clinical model of breast cancer incidence and progression is drawn from a large-scale, coordinated project funded by the National Cancer Institute (NCI) involving seven different research groups (Clarke et al. 2006); since there is naturally some uncertainty about the underlying model, we confirm that our main findings are not sensitive to a range of alternative assumptions about the onset and distribution of cancer type by age.

The estimates from our model indicate that women who would select into mammograms in the absence of the recommendation ("always-takers") have much higher rates of both in-situ and invasive cancer than the general population. We refer to this as "positive selection" into mammograms (positive with respect to cancer incidence). However, our estimates indicate that the women who select into mammograms due to the recommendation ("compliers") are much less likely to have invasive cancer—and are no more likely to have in-situ cancer—than women who do not select into mammograms ("never-takers"). The relative degree of selection pre- and post- the age-40 recommendation is identified directly from our data; the clinical model of underlying cancer incidence is needed to assess whether the observed selection either pre- or post-age 40 is positive with respect to the underlying population, whose cancer incidence is not directly observed.

We apply our model and its estimates to illustrate how the nature of selection in response to the recommendation affects the impact of the recommendation. Specifically, we estimate that shifting the recommendation from age 40 to age 45 results in more than three times as many deaths—at similar cost savings—if we assume that compliers with the recommendation are randomly drawn from the population rather than drawn based on the estimated selection patterns. We view this as a particularly instructive

[3]Because the context is naturally quite different, one might expect selection into compliance with RCT assignment to be different than selection into compliance with a recommendation. Indeed, never-takers comprise only 5% of the population in the context of Kowalski (2019), but are approximately 60% in ours.

counterfactual, since assuming that the women who respond are randomly drawn from the population is conceptually similar to using estimates of the impact of mammography from randomized experiments (with full compliance). Because in practice those who respond to the recommendation have a much lower rate of invasive cancer than the underlying population, the mortality cost of moving the recommendation to age 45 is lower than under random selection. Conversely, our model also illustrates that if it were feasible to target the recommendations to those with higher rates of cancer, the mortality cost of moving the recommendation from age 40 to 45 could be substantially larger than even the random selection assumption would imply. This is consistent with recent interest in reducing over-diagnosis by developing targeted, precision screening for women at higher risk (Elmore 2016; Esserman, Shieh, and Thompson 2009).

Our paper relates to several distinct literatures. Most narrowly, it speaks to the large body of work on mammograms, which we describe in the next section. But beyond the specific application of mammograms, it speaks to a broader health policy debate about whether and when to recommend medical screening tests (e.g., Welch, Schwartz, and Woloshin 2011). A central challenge that has limited empirical research on this topic is that—in the datasets typically available to researchers—the testing decision is observed but the outcome of the test is not. An attractive feature of our setting is that the outcome of the test (i.e. cancer incidence and type of cancer) is measurable both in claims data and in registry data. In this sense our analysis is similar in spirit to Abaluck et al. (2016), who are able to measure the outcome of imaging tests for pulmonary embolism in claims data, which they use to investigate whether and when that imaging test is being "overused." Both our paper and Abaluck et al. (2016) share a common feature with the racial profiling literature on stop and frisks (Anwar and Fang 2006; Persico 2009): the object of interest is only observed conditional on an action. This raises an empirical challenge for analyzing how the action (in our case, screening) relates to the underlying object of interest (in our case, the underlying incidence of cancer and cancer types). In our setting, we overcome this empirical challenge by combining two insights. First, the recommendation at age 40 serves as an exogenous source of variation in the screening rate, allowing us to estimate the cancer type of the marginal person affected by the recommendation. Second, the clinical oncology model of cancer incidence and growth allows us to use the observed moments (namely, outcomes conditional on screening under different regimes) to model outcomes under counterfactual regimes.

More broadly, our paper speaks to the value of complementing reduced-form estimates of causal effects with economic models of behavior, and particularly of selection. Reduced-form methods—both quasi-experimental and randomized experiments—aim to estimate causal effects by shutting down any endogenous choices. In practice, however, most policies involve an element of choice, so that the ultimate impact of the policy depends not only on the distribution of causal treatment effects but also on which women select into treatment. In this sense, our paper relates broadly to the literature on Roy selection, or selection on gains. In the health care context specifically, Einav et al. (2013) emphasize that the impact on health care spending of offering a high-deductible health insurance plan may be very different than what would be estimated from random assignment of high-deductible plans across individuals, because the types of people who choose high-deductible plans can have very different health care utilization responses to cost sharing than a typical individual. Our analysis speaks to a similar issue, in the context of evaluating recommendations for disease screening.

The rest of the paper proceeds as follows. Section I summarizes the relevant institutional details of our empirical context (breast cancer and mammography), and describes the existing evidence regarding the effect of mammograms and of various policy interventions that are designed to increase mammography rates. Section II describes our data and presents descriptive results. Section III presents our model of mammogram choice and describes how we estimate it using the observed descriptive patterns together with a clinical oncology model. Section IV presents the model estimates and discusses their implications for the impact of

changing the recommended age of beginning mammography under both observed and counterfactual selection patterns. The last section concludes by using our findings to speculate about possible policy implications more broadly.

I. Empirical context

A. Breast cancer

The earliest stages of breast cancer typically produce no symptoms and are not detectable in the absence of screening technologies.[4] As breast cancer progresses, it can spread within the breast, to adjacent tissues, to adjacent lymph nodes, and to distant organs (known as metastases). In clinical settings, tumors are classified according to the size of the tumor, the extent to which it has spread to lymph nodes, and whether it has metastasized. Public health research typically relies on a standardized classification—namely, the SEER classification system—which includes four stages: in-situ, local, regional, and distant; the last three stages are collectively referred to as "invasive" tumors.

Our analysis focuses on the distinction between in-situ and invasive tumors, a distinction that has been a key focus of the policy debate around mammography recommendations. In-situ refers to abnormal cells that have not invaded nearby tissues, instead remaining confined to the ducts or glands in which they originated. Some but not all in-situ tumors will become invasive. Expected survival time varies greatly by stage at diagnosis: women who are diagnosed with localized breast cancer are 99% as likely as cancer-free women to survive to 5 years after diagnosis, compared to 85% for regional breast cancer, and 27% for distant-stage breast cancer.[5] Within a stage, survival also varies with tumor size. For example, among women with regional disease, 5-year survival (again, relative to comparable cancer-free women) is 95% for tumors smaller than 2 centimeters in diameter, 85% for tumors of 2-5 centimeters, and 72% for tumors greater than 5 centimeters.[6]

B. Mammography

Asymptomatic breast cancer can be detected by a mammogram, which is a low-dose X-ray procedure that allows visualization of the internal structure of the breast. If an abnormality is detected on a routine screening mammogram, the woman is typically called back in for a diagnostic mammogram and—if needed—a confirmatory biopsy (Cutler 2008; Hubbard et al. 2011). Once a diagnosis has been confirmed, the woman may undergo surgery to remove the tumor, in addition to other treatments which aim to reduce the risk of recurrence, such as radiation therapy, chemotherapy, hormone therapy, and/or targeted therapy.

Mammography is based on the theory of early detection of invasive cancer, rather than detection and removal of precancerous lesions (Humphrey et al. 2002). The efficacy of mammography is the subject of considerable debate. Mechanically, mammography is most beneficial if machines can detect tumors in their earliest stages, and if tumors (on average) rapidly become more difficult to treat the longer they go undetected. The benefits from mammography will be lower if a tumor is slow to advance from stage to stage, if mortality when treatment begins at a later stage is similar to when tumors are treated earlier, or if mammogram machines are unlikely to correctly identify tumors. In practice, because most women diagnosed with breast cancer are treated immediately upon detection, there is little information about the

[4] Unless otherwise noted, the discussion in this section draws from the American Cancer Society (2017a).

[5] These tabulations are drawn from US SEER cancer registry data from 2007-2013, as in American Cancer Society (2017a).

[6] These tabulations are drawn from US SEER cancer registry data from 2000-2014, as in American Cancer Society (2017a).

natural history of breast cancer tumors, making it difficult to know how an individual tumor would have progressed had it not been treated (Zahl, Maehlen, and Welch 2008). This complicates attempts to quantify the benefits of mammography.

In principle, the major potential health benefit of mammography is reduced mortality. However, in practice, randomized trials of the impact of mammograms on mortality have documented mixed results (Habbema et al. 1986; Alexander et al. 1999; Miller et al. 2000, 2002; Nyström et al. 2002; Bjurstam et al. 2003; Moss et al. 2006). There have been nine trials in total, with the first one dating back to the 1960s (Welch and Black 2010). Their estimates of relative risk reduction in breast-cancer mortality due to invitation to mammography range from 0% to 31% (Welch and Passow 2014), but many of these studies have lacked the statistical power to separately determine effects in different age groups (Humphrey et al. 2002). In particular, while most studies indicate that mammography reduces mortality among average-risk women over age 50, recent trials specifically designed to study mammography in younger women (aged 40-49) have estimated statistically insignificant reductions in breast-cancer mortality in this age group (Bjurstam et al. 2003; Moss et al. 2006).

The potential costs of mammography include financial, physical, and psychological costs. These costs arise from the initial screening, the frequent finding of false positives, and the treatment of cancers that would not have become clinically relevant in a woman's lifetime (often referred to as "over-diagnosis") (Jørgensen and Gøtzsche 2009). Some of these costs, such as the financial cost of a screening, are easy to quantify, while others are much more difficult to estimate. Estimates of the rate of over-diagnosis of breast cancer (from both observational work and inferences from randomized control trials) range from less than 5% to more than 50% of diagnosed breast cancers (Zackrisson et al. 2006; Jørgensen and Gøtzsche 2009; Bleyer and Welch 2012; Oeffinger et al. 2015; Harding et al. 2015; Welch et al. 2016; Jørgensen et al. 2017).[7]

C. Age-specific mammogram recommendation and its impact

Several studies have combined the existing evidence to quantify the costs and benefits of mammograms (e.g., Welch and Passow 2014; Ong and Mandl 2015). For example, Welch and Passow (2014) estimate that for every 1,000 women aged 40-49 who undergo annual mammography for 10 years, 0.1-1.6 women will avoid dying from breast cancer, while 510-690 will have at least one false-positive result and up to 11 women will be over-diagnosed and (unnecessarily) treated. As the estimates of the costs and benefits of mammography have evolved, so have the recommendations by medical associations regarding which groups of women should receive mammograms, and how often.

In the 1980s, following the first randomized trials of routine mammography, the National Institutes of Health (NIH), the National Cancer Institute (NCI), and eleven other health care organizations issued recommendations for routine screenings of women over age 40 (Kolata 2009). These recommendations became the subject of controversy over time as more trials were published, and the US federal government subsequently reconsidered its position. In 1997, an NIH panel concluded that there was insufficient evidence to recommend routine screening for women in their 40s, a finding that was controversial (one radiologist described the finding as a "death sentence" for women (Taubes 1997)). After public pressure, the Senate encouraged an advisory board to reject that conclusion (Kolata 2009). In 2009, following the publication of experimental data that failed to show statistically significant mortality benefits of mammograms for women in their 40s, the US Preventive Services Task Force (USPSTF) recommended that women begin screening at age 50. Again, this conclusion generated backlash from patient advocacy groups like the American Cancer

[7]Selection into screening potentially (partially) explains the phenomenon of over-diagnosis, since it results in more diagnoses of low-risk tumors.

Society, which at the time recommended annual screening for women aged 40 and above (American Cancer Society 2018).[8] This negative reaction was exacerbated by fears that the Affordable Care Act (ACA, then being drafted) would allow insurers to refuse to cover mammograms for younger women. The USPSTF stood by its recommendation, but a poll found that 84% of women aged 35-49 did not plan to follow the new recommendations, and the ACA was modified to mandate that insurers reimburse mammograms for women aged 40 and over (Saad 2009). Although in the last few years most patient advocacy organizations have begun to moderate their stances, the question of whether mammography should be recommended in the 40-49 age group remains controversial.

Importantly, both the academic literature and the policy debate over the costs and benefits of mammograms have primarily focused on the average impacts of mammograms at specific ages. For example, Welch and Passow (2014) extrapolate results from mammography RCTs to the entire population without considering selection effects. In contrast, our focus is on the characteristics of women whose decision to get a mammogram is influenced by the mammogram recommendation, and how their underlying cancer incidence and characteristics may differ from that of a randomly-selected woman in the population.

Several papers have examined the mammogram response to recommendations (Kadiyala and Strumpf 2011, 2016; Jacobson and Kadiyala 2017). Most closely related to our work on the selected response to mammogram recommendations is Kadiyala and Strumpf (2016), who document a sharp increase in self-reported mammograms at age 40 and estimate that most of the "newly detected" cancers are early-stage cancers. Also closely related is the work of Kim and Lee (2017) and Bitler and Carpenter (2016), who document that women who elect to receive mammograms in response to price reductions are in better health than those who get the mammogram even without the price reduction or those who don't get the mammogram even with the price reduction. Finally, Kowalski (2019) shows that the compliers in a Canadian mammography RCT are healthier on both cancer dimensions (i.e., rates of breast cancer) and non-cancer dimensions (e.g. body mass and smoking) than the never-takers.

II. Data and descriptive patterns

A. Data and variable construction

Our analysis of mammogram choices and outcomes focuses on women aged 35-50 and draws on two primary data sources. The first is claim-level data provided by the Health Care Cost Institute (HCCI), consisting of all claims paid by three large commercial insurers (Aetna, Humana, and UnitedHealthcare) from January 2008 through December 2012. Together, these three insurers represented about one-quarter of individuals under age 65 with commercial insurance (HCCI 2012). The data capture the billing-related information contained in the claims that these insurers pay out to medical providers; this includes the exact date and purpose of each claim, as well as the amount paid by the insurer and the amount owed out of pocket. The data also include a (masked) person identifier as well as the individual's birth year and gender.

The claim-level information in the HCCI data allow us to construct variables measuring whether an individual had a screening mammogram,[9] whether the result was positive or negative, and whether a positive result was a true positive or false positive. Our coding of screening mammograms (hereafter "mammograms")—as well as their outcomes—broadly follows the approach of Segel, Balkrishnan, and Hirth

[8]The American Cancer Society currently recommends annual screening for women between ages 45-54 and screening every 2 years for women 55 years and older (American Cancer Society 2018).

[9]A "screening mammogram" is a routine test that is conceptually different – and coded differently in the data – from a "diagnostic mammogram," which would typically follow the emergence of a possible breast cancer symptom (such as a positive screening mammogram).

(2017), which we cross-validated using Medicare claims data linked to cancer registry data (see Appendix A for more details).

The complete HCCI data contain about 28.7 million privately-insured women aged 25-64, and over 70 million woman-years. We limit the data to woman-years aged 35-50 who are covered continuously for at least three years between January 2008 and December 2012; we keep all the years of coverage except the first and last (since for every woman-year we need to observe the previous year to define screening mammograms and the subsequent year to measure outcomes). This results in about 7.4 million woman-years, and 3.7 million distinct women over the years from January 2009 to December of 2011.

The primary drawback of the HCCI data is that we are not able to observe information on a breast cancer diagnosis beyond its detection. To overcome this limitation of the HCCI data, we also analyze the National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results (SEER) database. This is an administrative, patient-level cancer registry of all cancer diagnoses in 13 US states, covering about one quarter of the US population (SEER 2019). We analyze all the breast cancer diagnoses in the data between 2000 and 2014 for women aged 35-50 at the time of diagnosis; this covers about 212,000 diagnoses. All cancer diagnoses are required to be reported, with data collected directly from the cancer patients' medical records at the time of diagnosis (rather than self reports).[10] For each diagnosed cancer, the SEER data contain information about the size and stage of each tumor at diagnosis. They also contain basic demographics for the patient including age at time of diagnosis, race, and insurance coverage, as well as subsequent mortality information through December 2013.

In our HCCI sample, the average woman's age is 43 and 27% of woman-years are under 40. In the SEER data, because cancer risk increases with age, the average age at diagnosis is a bit higher (44.6) and only 13% of the SEER diagnoses occur in women under 40. In SEER, where we can observe race, slightly over three-quarters of the sample is white. And unlike the HCCI data where, by construction, everyone is privately insured, in the SEER data only 84% are privately insured, while 13% are on Medicaid.

Table 1 documents mammogram rates and test results in the HCCI data. About 30% of woman-years are associated with a mammogram. The vast majority (89.6%) of mammograms are negative, and another 9.7% are false positives. Only 0.7% are true positives. Among all woman-years with a mammogram, total (insurer plus out-of-pocket) health care spending in the 12 months starting from (and including) the mammogram averages $5,000; while it is slightly higher (by about $1,000) for those with a false positive, it is dramatically higher for those with true positives, averaging almost $50,000. Out-of-pocket spending in the 12 months post-mammogram is about $2,800 for women with a positive mammogram, compared to $715 for women with a negative mammogram and $950 for women with a false positive.

The SEER data provide more information on tumor stage and tumor size for the 212,000 true positives (i.e. diagnoses) we observe. Just over 15% are in-situ; the rest are invasive. Of the invasive, about 57% are localized, 38% are regional, and the remaining 5% are distant.

B. Mammograms and outcomes, by age

Figure 1 shows the age profile of annual mammogram rates in the HCCI data. Because we observe birth year, the mammogram rate at age, say, 40 is the share of women who got a mammogram in the year they turned 40. Between ages 39 and 41, the mammogram rate jumps by over 25 percentage points, from 8.9% to 35.2%. This pronounced jump in mammogram rates at age 40 has been previously documented in self-reported data

[10] See https://seer.cancer.gov/manuals/2018/SPCSM_2018_maindoc.pdf for more information. SEER registries are required to collect data on persons who are diagnosed with cancer and who, at the time of diagnosis, are residents of the geographic area covered by the SEER registry.

(Kadiyala and Strumpf 2011, 2016).[11] One might be concerned that the existence of a recommendation for mammograms at age 40 could bias upward self-reports at that age. However, our analysis, which uses claims data, confirms a real change in mammogram behavior at 40. Indeed, as we show in Appendix Figure A.1, the increase in mammogram rates that we estimate at age 40 in the HCCI data is very similar to what we estimate using self-reported data (from the Behavioral Risk Factor Surveillance System Survey, or BRFSS), although—consistent with prior work (Blustein 1995; Cronin et al. 2009)—we estimate lower mammogram rates at every age in claims data compared to self-reported data.

We examine the outcomes of these mammograms—negative, false positive, and true positive—by age in the HCCI data. As shown in Appendix Figure A.2, the vast majority (85 to 90 percent) of mammograms are negative, and almost all of the remainder are false positives; spending is much higher for true positives than false positives and negatives.

Figure 2a shows the share of mammograms that are true positive and false positive by age. Between ages 39 and 41, the share of true positives falls by one-third (from 0.84% to 0.56%). This indicates that the marginal women who choose to have a mammogram because of the screening recommendation at age 40 (i.e. "compliers") have lower underlying rates of cancer (i.e. true positive diagnoses) than those who choose to get screened at younger ages before the recommendation kicks in ("always-takers").

The share of mammograms that are false positive is generally declining smoothly in age because the probability of a false positive is higher for women with denser breast tissue, and density generally decreases with age (Susan G. Komen Foundation 2018). The exception is a small "spike" in false positives around age 40; this likely is attributable to the fact that the probability of a false positive mammogram is highest for a woman's first mammogram (American Cancer Society 2017b). Note, however, that while the share of mammograms that are false positive is trending fairly smoothly in age, the share of women experiencing a false positive rises considerably at age 40, since there is a 25-percentage-point increase in the share of women who have a mammogram. This is shown in Figure 2b: the share of women experiencing a false positive mammogram quadruples at age 40, from about 10 to 40 per thousand women.

Figure 3a documents the age profile of tumor type among all diagnoses in the SEER data. Between ages 39 and 41, the share of detected tumors that are in-situ (as opposed to invasive) rises by 6 percentage points, from 11.6 percent to 17.9 percent; this is consistent with prior findings from Kadiyala and Strumpf (2016). The average size of a detected tumor falls by over 10 percent, from 27.3mm at age 39 to 24.4mm at age 41, although the pattern is less dramatic since detected tumor size is also falling (albeit less rapidly) at earlier ages.

Finally, Figure 3b documents 5-year mortality post-diagnosis in the SEER data by age of diagnosis, separately for tumors initially diagnosed as in-situ and invasive tumors. Mortality is almost three times higher for invasive tumors compared to in-situ tumors. For example, at age 40, the five-year mortality rate is 16.2% for invasive tumors compared to 4.5% for in-situ tumors. However, the mortality rate is roughly flat by age within tumor type.

C. Who responds to the recommendation?

The preceding descriptive results from both the HCCI and SEER data suggest that the women brought into screening by the recommendation at age 40 have a lower cancer disease burden than those who sought screening prior to the age-40 recommendation. This manifests in lower rates of cancer, detection of cancer

[11] Our data span the time period when the 2009 US Preventive Services Task Force changed its recommendation for routine mammograms to begin at age 50 rather than at age 40. Past analyses, such as Block et al. (2013), have documented that this appears to have had little affect on women's mammography behavior, which is not surprising given the substantial public controversy over this recommendation change.

at earlier stages, and smaller tumors conditional on cancer detection among compliers compared to always-takers.

Naturally, we are also interested in comparing compliers to never-takers: those who do not get screened even after the age-40 recommendation is in effect. Since the cancer status of women who do not get screened is inherently difficult (or impossible) to observe, we will draw on a clinical model of breast cancer incidence and progression to estimate the cancer profile of never-takers. Before turning to this exercise in the next section, we can use the available data to compare compliers and never-takers on various non-cancer characteristics.

Specifically, we use the discrete onset of the recommendation at age 40 in a regression discontinuity framework to implement the Abadie (2002, 2003) approach to characterizing compliers and never-takers. Figure 4 shows the results. The left-hand panel compares various characteristics of compliers and never-takers; for completeness, the right-hand panel compares compliers to always-takers. The top panel examines preventive health behaviors and prior health care use in the HCCI data. The bottom two panels examine insured women in the BRFSS data; these data allows us to observe additional health behaviors and demographic characteristics. Appendix B contains more detail on the estimation approach and also shows the average characteristics of the population and the subset who receive a mammogram, by age.

Overall, Figure 4 suggests that women who receive a mammogram as a result of the recommendation are more likely to comply with other recommended preventive care than women who do not get a mammogram even in the presence of the recommendation. In particular, both data sets indicate that compliers are more likely to get flu shots and Papanicolaou tests (also known as Pap tests, which are used to screen for cervical cancer) than never-takers. The HCCI data also indicate that compliers have lower health care spending and have fewer emergency room visits than never-takers. These results are consistent with Oster (2020)'s finding that when a health behavior is recommended, those who take it up also tend to exhibit other positive health behaviors. The results are also broadly consistent with related patterns reported by Kowalski (2019) in the context of selection into participation in clinical trials. Interestingly, however, we find no evidence of pronounced differences between compliers and never-takers on non-healthcare dimensions; they look similar on other health behaviors (such as seat belt use and alcohol consumption) as well as on basic demographics.

III. Model and estimation

The empirical patterns documented in the preceding section indicate that the women who respond to the mammogram recommendation have a lower incidence of cancer than those who seek mammograms in the absence of a recommendation. To evaluate the implications of this selection for alternative, counterfactual timings of the screening recommendation (such as at age 45 instead of age 40), we write down a stylized model of mammogram decision making. We then estimate this model using the observed patterns shown in Section II combined with a clinical oncology model of the underlying cancer incidence in the population and tumor evolution in the absence of detection. The clinical oncology model provides the (hitherto absent) crucial information on the cancer disease burden of women who respond to the mammogram recommendation compared to women who do not. Naturally, we explore sensitivity to alternative clinical assumptions.

A. A descriptive model of mammogram choice

Consider a woman i in a given year she is observed in the data.[12] We model the annual decision of whether or not to have a mammogram; annual decision frequency seems natural given that mammogram screening

[12]We observe women for one, two, or three years. As discussed below, this is a static model, which does not use the panel dimension, so we essentially treat the entire data as a cross-section of woman-years, each denoted by i.

tends not to be done more frequently than once a year. Absent any recommendation to do so, we assume that the "organic" decision to have a mammogram follows a simple probit, so that

$$\Pr\left(m_i^o = 1\right) = \Pr\left(\alpha^o + \gamma^o a_i + \delta^o_{in-situ} I(c_i^{in-situ}) + \delta^o_{invasive} I(c_i^{invasive}) + \varepsilon_i^o > 0\right), \tag{1.1}$$

where m_i^o is an indicator for whether woman i had a mammogram in that observed year, a_i is woman i's age that year, $c_i = \{c_i^{in-situ}, c_i^{invasive}\}$ describes woman i's undiagnosed cancer status that year, and ε_i^o is a (standard) Normally distributed error term. Following our discussion in Section II, our baseline specification summarizes cancer status c_i with two indicator variables, one that indicates an in-situ tumor and another that indicates an invasive tumor; the omitted category is no cancer.

If it is recommended that woman i obtain a mammogram, we model her response to the recommendation as a second, subsequent decision that is taken within the same year. That is, if a woman has already decided to have a mammogram "organically" based on equation (1), a recommendation has no additional impact. But for women who decided not to have a mammogram organically (that is, $m_i^o = 0$), a second decision point arises due to the recommendation, and we model this second decision point in a similar fashion, except that the parameters are allowed to be different:

$$\Pr\left(m_i^r = 1 | m_i^o = 0\right) = \Pr\left(\alpha^r + \gamma^r a_i + \delta^r_{in-situ} I(c_i^{in-situ}) + \delta^r_{invasive} I(c_i^{invasive}) + \varepsilon_i^r > 0\right), \tag{1.2}$$

where ε_i^r is a (standard) Normally distributed error term, drawn independently from ε_i^o.[13] This model assumes that the impact of the recommendation is (weakly) monotone for all women. For each woman, it only increases the probability that she has a mammogram, a feature that seems (to us) natural.[14]

Since we do not directly observe whether a mammogram was taken for organic reasons or in response to a recommendation, the probability that woman i obtains a mammogram in the year she is observed is given by

$$\Pr\left(m_i = 1\right) = \begin{cases} \Pr\left(m_i^o = 1\right) & \text{if not recommended} \\ \Pr\left(m_i^o = 1\right) + \Pr\left(m_i^r = 1 | m_i^o = 0\right) \Pr\left(m_i^o = 0\right) & \text{if recommended} \end{cases}$$

We use the model's results to quantify the degree of selection into mammograms in the presence and absence of a recommendation, and to examine how the nature of this selection affects the impact of recommendations. To do so, we use the model estimates to predict mammogram rates and mammogram outcomes under the current recommendation to begin mammograms at age 40 as well as under a counterfactual recommendation to begin at age 45. Consistent with our focus on selection, we also examine how alternative, counterfactual selection into mammograms in response to the recommendation would change the impact of changing the recommended age of beginning mammography from 40 to 45.

Discussion. Importantly, this is a descriptive, or statistical model of mammogram choice, rather than a behavioral one. This is most apparent from the fact that we use the cancer status c_i as an explanatory variable, when naturally this cancer status is unknown by undiagnosed women. Cancer status c_i is also unobserved by the econometrician; we describe below the clinical model of tumor evolution which we use to "fill in" these missing data, thus essentially integrating over the population distribution of this cancer status component.

We take this modeling approach for several reasons. First, many of the outcomes in this setting are

[13] While this independence assumption may appear restrictive, note that equation (2) only applies to those women who elected not to obtain an "organic" mammogram. It is therefore effectively restricted to women with "low enough" ε_i^o's, so that much of the potential correlation is already conditioned out.

[14] That is, as in the analysis of Section II.C, we assume that there are no defiers. As will become clear later, other than appearing a natural assumption to us, it also simplifies the intuition of how counterfactual recommendation policies play out.

difficult to assess or monetize, e.g. the stress and anxiety associated with false-positive test results or the non-monetary costs associated with the breast cancer treatment (even if successful). This makes it difficult to translate the rich set of outcomes into a single metric of utility. Second, our key focus is on the impact of the recommendation policy. With a perfectly informed population of women, recommendations should have no impact, yet the data in Section II show a clear increase in the mammogram rate in response to the age 40 recommendation. We could try to attribute this recommendation-induced increase in mammogram rate to improved information, but this would require us to make assumptions about what type of information is being revealed and how, or why women did not have such information to begin with. We prefer instead to remain agnostic about the behavioral channel by which the recommendation affects screening rates. Finally, a descriptive model of decision making does not require us to try to reconcile observed patterns of decisions with optimal behavior, or model deviations from optimality. The drawback is, of course, that we will not be able to engage with other policy changes or with the impact of changes in the recommendation policy on individual welfare directly, but rather will only evaluate changes in recommendation policies through their effect on observed outcomes.

Another key feature of our setup is that we model the mammogram decision to be a static—and perhaps naive—one. The decision is static in the sense that we assume that women do not take into account, for example, the time elapsed since their most recent mammogram (if any).[15] The decision is naive in the sense that we assume that women, when deciding to get a mammogram or not, do not explicitly take into account their propensity to get a mammogram in future years. This assumption seems not unrealistic, and simplifies the model. This assumption is particularly important in the context of our counterfactual exercise, which holds the estimated model as given while we change the age at which it is recommended to begin mammography. Specifically, in considering the changes that occur when the mammogram recommendation begins at age 45 instead of 40, our static model assumes that this would have no impact on women aged 39 or younger. In a dynamic model with forward-looking agents, however, it could increase the propensity of women under age 40 to get a mammogram. Our current model could in principle capture such dynamics implicitly by allowing serial correlation in ε_i^o and in ε_i^r. However, because we have a relatively short panel, and because we only use age to match the two main data sets, it would be hard to identify such a serial correlation structure. Consistent with this being a fairly inconsequential assumption, Figure 2 shows very low rates of pre-recommendation mammograms, and no evidence that mammogram rates decline in the year or two that are right before age 40 (when forward-looking women might anticipate their future mammogram).

B. Implementation

A clinical model of tumor appearance and evolution. To complete the empirical specification, we specify a clinical oncology model of tumor appearance and tumor evolution. The oncology model has two important roles in our analysis, one for estimation and another for our counterfactual exercises. For estimation, the key role of the oncology model is that it allows us to "impute" cancer status for the "never-takers," i.e., the women who do not get screened even when it is recommended. This clinical model delivers two key elements. First, it produces the underlying incidence of cancer (and cancer type) by age. This cannot be directly observed in data since cancer incidence is only observed conditional on screening. Intuitively, since we observe the rate of cancer among those who get screened and the share of women who get screened,

[15] While restrictive, there is no strong evidence of such dynamic patterns in the data. We only have a short panel of at most three years for each woman, so it is difficult to apply any formal statistical testing. However, conditional on having two mammograms during the three years of mammogram claims we observe (2009-2011), the frequency of getting a mammogram "every other year" (that is, getting mammograms in 2009 and 2011 but not in 2010) is not more likely than getting a mammogram in consecutive years (34%, relative to 39% for 2009 and 2010, and 27% for 2010 and 2011).

then, with the estimate of the overall rate of cancer from the clinical model, we can deduce the rate of cancer in the unscreened population. Second, the clinical model provides (counterfactual) predictions for the rate at which tumors would progress in the absence of detection and treatment (the so-called "natural history" of the tumor). Since breast cancer is usually treated once diagnosed, rather than being monitored without treatment, it is difficult (perhaps impossible) to directly estimate the natural history of tumors from existing data. This latter element is particularly important for our counterfactual exercises, in which the effect of different selection patterns depends on the share of cancer cases that get diagnosed, as well as how early tumors are found. In order to assess how clinically important early diagnosis is (e.g., in its effect on mortality), a model of tumor evolution is needed.

For the clinical model, we draw on an active literature creating clinical/biological models of cancer arrival and growth. Specifically, we draw on the work of the Cancer Intervention and Surveillance Modeling Network (CISNET) project funded by the National Cancer Institute to analyze the role of mammography in contributing to breast cancer mortality reductions over the last quarter of the 20th century. As part of this effort, seven different groups[16] developed models of breast cancer incidence and progression (Clarke et al. 2006). For convenience, we focus on one of these models, the Erasmus model (Tan et al. 2006). As we discuss below, we also confirm that our main results are not sensitive to alternative specifications designed to produce markedly different estimates for the key objects (the underlying incidence of cancer and cancer types).

We briefly summarize the Erasmus model here; Appendix C describes the model in much more detail. Starting with a cancer-free population of 20-year-old women, the Erasmus model assumes that breast tumors appear at a given age-specific rate (that is increasing in age). When they appear, tumors are endowed with a given invasive potential and initial rate of growth, and then evolve accordingly over time with respect to those two characteristics. Tumors can either be invasive, leading to death of the women if not detected early enough, or be in-situ. In-situ tumors are not themselves harmful but may either transform into a harmful invasive tumor or remain benign. In some sense, a key issue in the debate over mammograms is the extent to which tumors that are detected early (e.g. in-situ tumors) would have become harmful if not detected or would have remained benign; Marmot et al. (2013) discusses how, depending on the method of analysis, a wide variety of estimates can be obtained when trying to answer this question. The Erasmus model further classifies tumors by whether or not they are detectable by screening, which in the case of invasive tumors depends on their size and in the case of in-situ tumors depends on their sub-type. Finally, the model assumes that beyond a certain size, invasive tumors are fatal.

The original Erasmus model was calibrated using a combination of Swedish trial data and US (SEER) population data. To better match the cancer incidence rates in the SEER data (birth cohorts 1950-1975), we introduce a proportional shifter of overall cancer incidence and calibrate this parameter on the SEER data. Appendix Figure A.6 shows the calibrated model's predictions—under the assumption of no screening—of the share of women with cancer at each age, and the share of existing cancers that are in-situ (rather than invasive) by age.

Estimation and identification. We estimate the model using method of moments. The observed moments we try to match are the mammogram screening rate at each age (Figure 1), the true positive rate at each age (Figure 2a), and the share of tumors at each age that are in-situ conditional on true positive (as in Figure 3a).[17] Because identification is primarily driven by the discontinuous change in screening rates at age 40,

[16]The composition of the CISNET consortium has changed over time, but the seven groups who produced models for the original publication in 2006 were affiliated with the Dana-Farber Cancer Center, Erasmus University Rotterdam, Georgetown University Medical Center, University of Texas M.D. Anderson Cancer Center, Stanford University, University of Rochester, and University of Wisconsin-Madison.

[17]Figure 3a shows the share of all diagnosed cancers (in the SEER data) that are in-situ, but the model produces a different

we weight more heavily moments that are closer to age 40 than moments that are associated with younger and older ages.[18]

To generate the corresponding model-generated moments, we simulate a panel of women starting at age 20, and use the clinical model described above to generate cancer incidence and tumor growth for each woman. We then apply our mammogram decision model, by age and recommendation status, to each simulated woman who is alive and has yet to be diagnosed with cancer. The simulated cohort allows us to see the fraction of women with a detectable (by mammogram) tumor at each age, and thus generate the mammogram rate, and the true positive rate (by cancer type) conditional on screening. As mentioned above, for cancer type, we distinguish only between in-situ and invasive tumors.

With this simulated population of women, an assumed value of parameters associated with the mammogram decisions with and without recommendation (equations (1) and (2)) and the observed policy recommendation (40 and above), the model generates an age-specific share of women who are screened, and the tumor characteristics (in-situ and invasive rates), conditional on getting screened. We then search for the parameters that minimize the (weighted) distance between these generated moments and the observed moments described above.

Although the model is static, it does have a dynamic element because we calculate the model-generated moments only for women who were not diagnosed with cancer in previous years, and for those who did not die (from breast cancer or other causes) prior to the given age. Specifically, because the mammogram decision applies to women who have yet to be diagnosed with cancer, fitting the model requires calculating the rate of cancer among the population who is eligible to be screened, which includes those who have currently undiagnosed cancer or no cancer, but does not include those who are dead or already diagnosed. Appendix D provides more detail on this and other aspects of the estimation.

For our counterfactual exercises, the estimates from the mammogram choice model—and the assumption that choices would be smooth in age through age 40 in the absence of the recommendation—allow us to predict mammogram decisions and outcomes under counterfactual scenarios. Crucially, the model estimates allow us to forecast the cancer characteristics of women who (counterfactually) do not get screened and whose cancer may therefore progress in the absence of diagnosis. The key parameters are δ^o and δ^r, which capture the nature of selection into mammogram screening. Positive selection (i.e. positive δ) implies that women with cancer (or with invasive vs. in-situ cancer) are more likely to get a mammogram than are woman without cancer. A negative δ implies the opposite. Both types of selection are plausible. Positive selection could arise, for example, if women with a greater risk of breast cancer (e.g. due to family history) are more likely to get a mammogram; negative selection could arise, for example, if women with certain underlying characteristics (e.g. risk aversion) are both more likely to get a mammogram and also more likely to avoid risk factors linked to breast cancer. Importantly, by allowing δ^o and δ^r to be different, the model allows for the nature of selection to be different for organic and recommendation-driven mammograms. Identification of these selection effects is driven by comparing the share of cancer in the population (which is "data" provided by the clinical oncology model) to the true positive mammogram rates. The extent to which this relationship changes discretely at age 40, when the recommendation kicks in, allows us to separately identify δ^o and δ^r.

metric: the share of screening mammogram-diagnosed cancers that are in-situ. Cancers that are clinically diagnosed are highly unlikely to be in-situ, so the SEER value likely underestimates the true value of share in-situ for screening mammogram-diagnosed cancers. Appendix D describes how we adjust the SEER moments to account for this.

[18]Specifically, the weight on moments associated with ages 39 and 41 is 10/11 of the weight on the age 40 moment, the weight on moments associated with ages 38 and 42 is 9/11 of the weight on the age 40 moment, and so on.

IV. The impact of alternative screening policies

A. Model fit and parameter estimates

Figure 5 presents the model fit to the key moments, which we view as quite reasonable. The parameter estimates are shown in Table 2. It may be easiest to see the implications of these parameters in the context of our counterfactual results, but one can already infer the general pattern by focusing on the four δ parameters, which indicate the extent of selection into mammogram. The two δ^o parameters are positive and relatively large, indicating strong positive selection into the "organic" decision to have a mammogram. For example, for the average woman-year in the sample (that is, using the distribution of ages in the sample), the estimated coefficients imply that the "organic" mammogram rates for women with either an in-situ or invasive tumor are much higher (0.30 and 0.57, respectively) relative to the "organic" mammogram rates for cancer-free women (0.20).

In contrast, the two δ^r parameters tell a different story. The estimates suggest that there is no differential selection into the "recommended" decision for women with in-situ tumors (relative to cancer-free women), and that essentially no woman with an invasive tumor selects into mammogram due to the recommendation. This result is driven by precisely the patterns in the data that identify these parameters, and which were presented in Figure 3a. Namely, conditional on diagnosis, the share of in-situ tumors rises sharply at age 40, so that virtually all the increase in detected cancers reflects in-situ tumors. As we show below, this pattern has a critical effect on our results, because women without cancer or with in-situ tumors—who constitute the primary incremental positive mammogram results—may not face drastic health implications if those tumors would instead be discovered several years later.

We note that the large confidence intervals on $\delta^o_{invasive}$ and $\delta^r_{invasive}$ reflect the fact that the estimates imply that virtually all women with invasive tumors who get screened do so organically, with essentially no women with invasive tumors getting screened in response to the recommendation; as a result, the likelihood function is fairly flat for high values of $\delta^o_{invasive}$ and low values of $\delta^r_{invasive}$. But for exactly the same reason, these imprecise estimates of the parameter have little impact on the counterfactual results, as reflected by the much tighter standard errors associated with the counterfactuals of interest reported in the next section.

B. Implications

We apply the estimated parameters from Table 2 to analyze outcomes under various counterfactual recommendations. For concreteness, we focus on outcomes under the current recommendation to begin mammograms at age 40 as well as under a counterfactual recommendation to begin at age 45. Our model is well suited for such a counterfactual exercise: we simply assume that mammogram decisions are based on the "organic" decision until age 45, and only at age 45 is there a second, recommendation-induced decision. Given the static nature of the model, mammogram rates will remain the same until age 40, and would be the same (conditional on cancer status) from age 45 and on, but will decrease for women aged 40-44 without a recommendation. We choose a counterfactual recommendation that begins at age 45 because this is not too far out of sample, and also in the range of realistic policy alternatives; Canada, for instance, recommends routine screening beginning at age 50 (Kadiyala and Strumpf 2011). Of course, such counterfactuals do require us to rely on our assumption of a linear age profile in order to predict outcomes for always-takers beyond age 40 in a counterfactual world in which the recommendation does not occur until age 45; while this strikes us as not unreasonable, given that the linear specification in age seems to fit the data well, it is of course an important (and untestable) assumption.

For both the age 40 and age 45 recommendations, we also examine how alternative, counterfactual

selection into mammograms in response to the recommendation would change the recommendation's impact. The main outcomes we generate under the various counterfactuals are age-specific mammogram rates, mammogram outcomes (specifically, negative, false positive, and true positive, as well as tumor type), total health care spending, and mortality. We do not attempt to quantify other potential consequences of a change in recommendation (such as the opportunity to use less invasive treatments for early-stage diagnoses, or increased anxiety from false positive results, which are more uncertain (Welch and Passow 2014)).

Throughout the counterfactual exercises, mammogram rates are generated directly from the parameter estimates in Table 2, and mammogram outcomes are generated based on the parameter estimates in Table 2 and the underlying incidence and natural history of breast cancer tumors from the Erasmus model. We also use the Erasmus model's parameters in order to map detection of tumors to subsequent mortality, allowing us to translate the estimated changes in detection into implied changes in mortality. Finally, we use the auxiliary data from Figure A.2b on how health care spending varies with age and mammogram outcomes to translate the estimated change in mammogram rates and mammogram outcomes into implied spending changes. Appendix E provides more details behind these counterfactual calculations.

Shifting the age of recommendation from 40 to 45. Table 3 shows the implications of shifting the recommendation from age 40 to age 45, given the estimated response to recommendations from Table 2. We focus on the implications for women ages 35-50.

Panel A summarizes the implications for screening and spending; Figure 6 shows how the age profile of screening and screening outcomes change with this counterfactual. Changing the recommended age from 40 to 45 reduces the average number of mammograms a woman receives between ages 35 and 50 from 4.7 to 3.8, an almost 20 percent decline. By design, all of the "lost" mammograms occur between ages 40 and 44. Naturally, the vast majority of these "lost" mammograms would have been negative (89.5%) or false positive (10.4%). Moving the recommendation to age 45 decreases the average number of false positives a woman experiences over ages 30-45 by 0.09. The fraction of true positive mammograms that are "lost" due to the later recommendation, while small in absolute number (0.0004 per woman), is not negligible, and it constitutes an approximately 6% reduction in the cancer detection rate. Of the "lost" true positives, however, all are in-situ since our estimates imply that the recommendation effectively induces no additional women with invasive cancer to get screened. Thus, any changes in mortality are due to in-situ tumors that go unscreened and later become invasive.

The last row of Panel A shows that changing the recommendation age to 45 reduces total health care spending over ages 35-50 per woman by about $320, or about half a percent. This reduction in spending arises from a combination of a level and composition effect. The dominant factor is naturally the decline in the overall number of mammograms. We estimate that women who have a mammogram in a given year are expected to spend approximately $570 more (on average, averaging over ages 40-44) over the subsequent 12 months relative to women with no mammograms, and that moving the recommendation age to 45 results in 0.9 fewer mammograms per woman. This would mechanically result in approximately $510 lower spending. The estimated spending reduction is lower ($320) because of selection. The "lost" mammograms are disproportionately negative or false positive, and the true positive mammogram results are associated with, by far, the highest expected subsequent spending (see Figure A.2b). True-positive mammograms account for a larger share of mammograms in the counterfactual scenario (0.53%, relative to 0.44% under the age-40 recommendation).

Panel B documents the implications of this counterfactual for health outcomes. The lower detection rate of cancers is associated with 5 more women per 100,000 who are dead by the age of 50; all of this increase in deaths comes from increased breast cancer mortality. The results thus suggest that, relative to an age-45 recommendation, an age-40 recommendation increases spending by about $32 million per 100,000 women

(during the ages of 35-50), and prevents about 5 additional deaths by age 50 per 100,000 women; the cost per life saved is thus about $6 million.

Naturally, these mortality implications are driven by the assumptions in the clinical oncology model, about which there is a range of views (Clarke et al. 2006; Welch and Passow 2014). In addition, our analysis considers only the costs in terms of health care spending, and does not consider the disutility of stress and anxiety created by false positives or additional medical care. For both reasons, our goal here is not to emphasize a specific estimate of the cost per life saved per se, but rather to examine whether and how this type of counterfactual policy exercise can be affected by the nature of selection into mammograms in response to the recommendation, a question we turn to in the next section.

Consequences of selection patterns in response to mammogram. Table 4 illustrates the importance of selection in response to the recommendation. To do so, Panel A replicates the results from Table 3, while Panels B and C contrast them with what the results would be under alternative selection responses to the recommendation. Under both alternative selection models, we maintain our estimated selection associated with the "organic" mammogram decision, but vary the nature of selection into mammograms in response to the recommendation. One case (Panel B) assumes no selection, which is conceptually consistent with the idea of using estimated mammogram treatment effects from randomized experiments to inform the recommendation policy (as in, for example, Welch and Passow 2014); in practice we do this by assuming that $\delta^r = 0$.[19] The other case (Panel C) assumes that selection in response to the recommendation is positive, and is the same as in the "organic" decision; we implement this counterfactual by assuming that δ^r is equal to our estimated δ^o.

In both counterfactual selection cases we consider, we adjust the model to maintain the same age-specific mammogram rates under a given recommendation regardless of the assumed selection, so that only the nature of selection changes; Appendix E provides more detail. By design, therefore, the mammogram rates (first row of each panel) remain almost the same across all three selection models,[20] and therefore the spending effect associated with each of these cases also remains almost identical (second row of each panel). In contrast, the importance of selection is shown in the third row of each panel: different patterns of selection affect the reduction in deaths from moving the recommendation to age 40 compared to age 45. For example, while our estimates that are based on observed selection imply that moving the recommendation from 45 to 40 saves 5 additional lives (by age 50) per 100,000 women, which corresponds to a cost of about $6.3 million per life saved, random selection would imply over three times as many lives saved (18 per 100,000), corresponding to a cost of about $1.9 million per life saved. At a more extreme case of selection, assuming that the strong positive selection associated with "organic" selection would also apply to the selection in response to the recommendation, would imply almost nine times as many lives saved (45 per 100,000 women), corresponding to a cost per life saved of about $0.86 million.

The qualitative results are intuitive. As selection associated with the recommendation is more negative (i.e. women who respond are less likely to have cancer), the recommendation for earlier mammograms is less effective in finding tumors that would have not been found otherwise or tumors that would otherwise be found only later. However, if the selection associated with the recommendation were very positive (i.e. women who respond are more likely to have cancer), an earlier recommendation would be more effective. Thus, out of the three selection scenarios considered, earlier recommendation is most beneficial if the selection

[19] Note that here we have in mind a conceptual randomized experiment with full compliance. Of course, in practice, full compliance is rare, and the complier population to the experiment is itself not random, although it may be differentially selected from the complier population to the recommendation. In a recent paper, Kowalski (2019) argues that in practice the women most likely to receive mammograms when encouraged to do so in a randomized clinical trial are healthier, and hence benefit less from mammograms.

[20] Although not seen in the table due to rounding, the mammogram rates are not exactly the same across the panels because the nature of selection leads to differential mortality (discussed below), which in turn (slightly) affects the set of women "eligible" for a screening mammogram.

response to the recommendation is the same as under "organic" selection, which was highly positive (Panel C). While it is not immediately clear how in practice to achieve such strong positive selection in response to the recommendation, this result suggests that better targeting of the recommended mammogram to women with higher a-priori risk of cancer could—if feasible—have dramatic effects on the mortality benefits from the recommendation.[21] The comparison between our estimated selection (panel A) and the "no selection" case (panel B) is an intermediate case. Because we estimate negative selection for invasive tumors, an earlier recommendation is more effective (i.e. more women with cancer would be screened) under random selection, and the cost per life saved is therefore lower.

Sensitivity. The data allow us to estimate characteristics of always-takers and compliers, and to see that compliers have a lower incidence of cancer than always-takers (see Figures 2a and 3a). However, our counterfactuals require us to also estimate the cancer status of never-takers, as well as how cancer would evolve if (counterfactually) screening occurred at a later age. For both of these endeavors, we relied heavily on the underlying natural history ("clinical") model of breast cancer. We therefore examine the sensitivity of our conclusions to changing key features of this model, such as the underlying incidence rate of cancer, the share of in-situ tumors that will become invasive if not treated, and the share of tumors that are non-malignant, i.e. have no potential to be invasive and therefore would never result in a breast cancer mortality.

This sensitivity analysis serves to highlight a point we have tried to emphasize throughout: the reader should not place much (or any) weight on our particular, quantitative estimates of the cost per life saved of recommending that mammography begin at 40 instead of at 45; these are quite sensitive to the assumptions underlying the clinical model. By contrast, the qualitative result we focus on—how the nature of the selection response to the recommendation affects any estimate of the impact of an earlier recommendation—is quite robust to alternative assumptions in the underlying clinical model. Appendix F discusses the specifics of how we implement the sensitivity analysis and presents the results in detail.

V. Summary and possible policy implications

The debate over whether and when to recommend screening for a particular disease involves a host of empirical and conceptual challenges with which the existing literature has grappled, including how to estimate the "health" return to early screening, how to measure non-health benefits or costs, and how to monetize all of these factors (Humphrey et al. 2002; Nelson et al. 2009; Marmot et al. 2013; Welch and Passow 2014; Ong and Mandl 2015). We make no pretense of "resolving" these issues. Instead, we suggest an additional important and largely overlooked factor that can—and should—be considered: the nature of selection in response to the recommendation.

We illustrate this point in the specific context of the (controversial) recommendation that women should begin regular mammogram screenings at age 40. We document that this recommendation is associated with a sharp (25 percentage point) increase in mammogram rates, and that those who respond to the recommendation have substantially lower rates of cancer incidence than those who choose to get mammograms in the absence of the recommendation (i.e. before age 40). Conditional on having cancer, women who respond to the recommendation also have lower rates of the more lethal invasive cancer, relative to the less lethal in-situ cancer. These data speak directly to the relative cancer risks of women who select mammograms in the absence and presence of a recommendation. To further assess how the cancer risk of those who

[21] The potential benefits of personalizing breast cancer screening recommendations have highlighted in the medical literature (e.g. Schousboe et al. 2011), and current breast cancer screening recommendations often differ across average-risk and high-risk women (where the latter is, e.g., women with a family history of breast cancer). But to the best of our knowledge our point about selection responses to recommendations has not been made previously. Our consistent selection model is one way of illustrating the potential gains from recommendation designs that affect take-up of mammograms based on unobservables.

select mammograms when recommended compares to those who do not select mammograms even when recommended, we draw on a clinical oncology model to estimate the underlying cancer incidence in the non-screened population (since this is not directly observed). These results suggest that those who choose mammograms in the absence of a recommendation have substantially higher rates of both invasive and in-situ cancer than women who do not get screened; women who choose mammograms in response to the recommendation have similar rates of in-situ cancer to unscreened women but much lower rates of invasive cancer than unscreened women.

To illustrate the potential consequences of these selection responses to recommendations, we write down a stylized model of the mammogram decision, which depends on age, cancer status, and recommendation. We estimate this model using the observed empirical patterns combined with the clinical oncology model, the latter of which provides both the underlying incidence of cancer and the (counterfactual) tumor evolution in the absence of detection. We then apply the model to assess the implications for spending and mortality of changing the recommended age for beginning mammograms from 40 to 45. The specific numbers that we estimate will naturally be sensitive to the modeling assumptions; moreover, our estimates do not attempt to measure all of the the potential impacts of mammograms, such as stress.

Our focus instead is on the consequences of the selection response to the recommendation, which our estimates suggest are non-trivial. Specifically, we consider the impact of moving the recommended age of beginning mammography from 45 to 40, and how this varies under alternative selection responses to the recommendation. We hold the change in mammogram rates (and consequently the cost increase) from changing the recommended age constant, and show that the mortality implications from earlier recommended mammograms vary markedly with selection patterns. For example, under the observed selection pattern, the number of lives saved by moving the recommendation from age 45 to 40 is less than a third of what it would be if those who responded to the recommendation were instead drawn at random from the population. This difference arises because we estimate that those who respond to the recommendation have much lower rates of invasive cancer. Conversely, our results also suggest that if it were feasible to target the recommendations to those with higher rates of cancer, shifting the recommendation from age 45 to 40 would save substantially more lives than either the observed selection patterns or random selection.

These findings suggest that the ongoing debates over whether and when to recommend screening for a disease should consider not only average costs and benefits from screening, but also the nature of selection associated with those who respond to the recommendation. They also suggest that future work exploring the impact of existing policy instruments or the design of potential new ones should consider not just aggregate impacts on mammography rates, but also the cancer incidence for compliers.

While our empirical focus has been on recommendations, these are of course only one part of a broader set of policy efforts that have been deployed or discussed for increasing disease screening. In the case of mammograms, another widely-used instrument has been lowering the financial costs of screenings. For example, in 1991 the federal government launched the National Breast and Cervical Cancer Early Detection Program to provide free screenings to women below 250 percent of the federal poverty line (Lee et al. 2014). In the same year, Medicare expanded its coverage to include bi-annual screening mammograms; subsequently, in 1998, Medicare expanded coverage further to include annual screening mammograms and to waive the deductible (O'Sullivan et al. 1997; Kelaher and Stellman 2000; Habermann et al. 2007). On the private insurance side, a number of states have mandated that insurance plans must cover mammography (Bitler and Carpenter 2016). Beyond these financial levers, there are also policy efforts to reduce non-financial barriers to mammograms. These include, for example, increasing ease of access to mammograms through programs such as mobile mammography clinics (Vang, Margolies, and Jandorf 2018), and outreach efforts designed to educate women about the benefits of mammograms and informing them of the services available

to them (Levano et al. 2014).

Related to these efforts, an existing literature has studied the impact of various policy instruments on mammography rates. It has found, for example, that lowering out-of-pocket financial costs increases mammogram rates (Kelaher and Stellman 2000; Habermann et al. 2007; Finkelstein et al. 2012; Fedewa et al. 2015; Mehta et al. 2015; Bitler and Carpenter 2016; Cooper et al. 2017; Kim and Lee 2017), while increasing the distance a woman must travel to get a mammogram decreases mammogram rates (Lu and Slusky 2016). Only a few of these studies have examined differential responses to the policy by underlying health characteristics. This existing work suggests that, like our findings on the response to guidelines, those who get mammograms in response to a lower price and those who comply with assignment to mammogram treatment in a clinical trial tend to be healthier than never-takers and always-takers (Bitler and Carpenter 2016; Kim and Lee 2017; Kowalski 2019). While one of course must be careful in generalizing too much from a few studies, our read of this existing literature is that these alternative interventions are not obviously better targeted than recommendations in terms of the compliers, at least in the context of mammograms.

The combined evidence therefore highlights the importance of trying to better target the existing instruments. This is challenging since underlying cancer incidence, tumor stage, and tumor size are not observable without screening. However, our descriptive analyses in Section II—comparing compliers to never-takers on a host of observable characteristics—suggest that never-takers are also less likely than compliers to engage in other recommended health behaviors, such as flu shots and Pap tests. This finding is consistent with the idea that those who comply with recommendations tend to exhibit other positive health behaviors (Oster 2020). It also suggests that coordinated efforts, which attempt to draw in women who otherwise would not engage in any preventive health behaviors, could be high-value. If we are willing to extrapolate our qualitative results from breast cancer to these related contexts, our findings suggest that trying to get such women to undertake a slew of recommended health behaviors might be well-targeted at reaching women at higher risk of not only breast cancer, but perhaps also cervical cancer and the flu.

Recent analyses by clinical researchers also suggest other observables that might be useful in targeting mammograms to higher-risk groups, instead of (or in conjunction with) age-based screening recommendations. For example, Evans et al. (2019) describe the results of a randomized trial that begins regular mammograms at age 34 for women with a mother or sister who has been diagnosed with breast cancer, and an ongoing trial is investigating the impact of risk-based screening relative to standard annual screening (Esserman et al. 2017). Motivated by such work, researchers have proposed that the recommended age of beginning mammography should be based on individual risk factors such as age of first birth, number of children, and breast density (Evans, Howell, and Howell 2020; Mukama et al. 2020). Our findings underscore the potential value of such targeting, given that compliance with the existing recommendation is only about one-third, and compliers appear to be disproportionately low-risk for cancer. They also suggest the importance of analyzing the impact of targeted instruments, not only for recommendations but also for price subsidies and other policy instruments.

More broadly, our findings suggest that considering and improving selection into screening is a first-order factor in an effective design and analysis of interventions to increase screenings. However, the extent to which we can generalize our findings in this paper, in the context of mammograms, to other types of screening and preventive medicine remains an open question. The controversy surrounding the recommendation that mammography start at age 40 may generate stronger selection than in other, less controversial settings (e.g., flu shots). Whether this is true or not is an important question that we leave for future work.

References

Abadie, Alberto. 2002. "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models." *Journal of the American Statistical Association* 97 (457): 284–92.

Abadie, Alberto. 2003. "Semiparametric Instrumental Variable Estimation of Treatment Response Models." *Journal of Econometrics* 113 (2): 231–63.

Abaluck, Jason, Leila Agha, Chris Kabrhel, Ali Raja, and Arjun Venkatesh. 2016. "The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care." *American Economic Review* 106 (12): 3730–64.

Alexander, F.E., T.J. Anderson, H.K. Brown, A.P.M. Forrest, W. Hepburn, A.E. Kirkpatrick, B.B. Muir, R.J. Prescott, and A. Smith. 1999. "14 Years of Follow-Up from the Edinburgh Randomised Trial of Breastcancer Screening." *Lancet* 353 (9168): 1903–08.

American Cancer Society. 2017a. "Breast Cancer Facts & Figures 2017-2018." American Cancer Society, Inc. https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2017-2018.pdf.

American Cancer Society. 2017b. "Limitations of Mammograms." https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection/mammograms/limitations-of-mammograms.html.

American Cancer Society. 2018. "History of ACS Recommendations for the Early Detection of Cancer in People Without Symptoms. https://www.cancer.org/health-care-professionals/american-cancer-society-prevention-early-detection-guidelines/overview/chronological-history-of-acs- recommendations.html.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–65.

Anwar, Shamena, and Hanming Fang. 2006. "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence." *American Economic Review* 96 (1): 127–51.

Berry, Donald A. 2013. "Breast Cancer Screening: Controversy of Impact." *Breast* 22: S73–76.

Bitler, Marianne P., and Christian S. Carpenter. 2016. "Health Insurance Mandates, Mammography, and Breast Cancer Diagnoses." *American Economic Journal: Economic Policy* 8 (3): 39–68.

Bjurstam, Nils, Lena Björneld, Jane Warwick, Evis Sala, Stephen W Duffy, Lennarth Nyström, Neil Walker, Erling Cahlin, Olof Eriksson, Lars-Olof Hafström, Halvard Lingaas, Jan Mattsson, Stellan Persson, Carl-Magnus Rudenstam, Håkan Salander, Johan Säve-Söderbergh, and Torkel Wahlin. 2003. "The Gothenburg Breast Screening Trial." Cancer, 97 (10): 2387–96.

Bleyer, Archie, and H. Gilbert Welch. 2012. "Effect of Three Decades of Screening Mammography on Breast-Cancer Incidence." *New England Journal of Medicine* 367 (21): 1998–2005.

Block, Lauren D., Marian P. Jarlenski, Albert W. Wu, and Wendy L. Bennett. 2013. "Mammography Use Among Women Ages 40–49 After the 2009 U.S. Preventive Services Task Force Recommendation." *Journal of General Internal Medicine* 28 (11): 1447–53.

Blustein, Jan. 1995. "Medicare Coverage, Supplemental Insurance, and the Use of Mammography by Older Women." *New England Journal of Medicine* 332 (17): 1138–43.

Brett, J., C. Bankhead, B. Henderson, E. Watson, and J. Austoker. 2005. The Psychological Impact of Mammographic Screening. A Systematic Review." *Psycho-Oncology* 14 (11): 917–38.

Clarke, Lauren D., Sylvia K. Plevritis, Rob Boer, Kathleen A. Cronin , and Eric J. Feuer. 2006. "A Comparative Review of CISNET Breast Models Used To Analyze U.S. Breast Cancer Incidence and Mortality Trends." *Journal of the National Cancer Institute, Monographs* (36): 96–105.

Cooper, Gregory S., Tzuyung Doug Kou, Avi Dor, Siran M. Koroukian, and Mark D. Schluchter. 2017. "Cancer Preventive Services, Socioeconomic Status, and the Affordable Care Act."

Cancer 123 (9): 1585–89.

Cronin, Kathleen A., Diana L. Miglioretti, Martin Krapcho, Binbing Yu, Berta M. Geller, Patricia A. Carney, Tracy Onega, Eric J. Feuer, Nancy Breen, and Rachel Ballard-Barbash. 2009. "CEBP Focus on Cancer Surveillance: Bias Associated with Self-Report of Prior Screening Mammography." *Cancer Epidemiology, Biomarkers, and Prevention* 18 (6): 1699–1705.

Cutler, David M. 2008. "Are We Finally Winning theWar on Cancer?" *Journal of Economic Perspectives* 22 (4): 3–26.

Einav, Liran, Amy Finkelstein, Abigail Ostriker, Tamar Oostrom, and Heidi Williams. 2020. "Screening and Selection: The Case of Mammograms." *American Economic Review* 110 (12): 3836–3870.

Einav, Liran, Amy Finkelstein, Stephen Ryan, Paul Schrimpf, and Mark R. Cullen. 2013. "Selection on Moral Hazard in Health Insurance." *American Economic Review* 103 (1): 178–219.

Elmore, Joann G. 2016. "Solving the Problem of Overdiagnosis." *New England Journal of Medicine* 375 (15): 1483–86.

Esserman, Laura J., Yiwey Shieh, and Ian Thompson. 2009. "Rethinking Screening for Breast Cancer and Prostate Cancer." *Journal of the American Medical Association* 302 (15): 1685–92.

Esserman, Laura J., and the WISDOM Study and Athena Investigators. 2017. "The WISDOM Study: Breaking the Deadlock in the Breast Cancer Screening Debate." *Breast Cancer* 3 (1): 1–7.

Evans, D. Gareth, Sacha J. Howell, and Anthony Howell. 2020. "New Evidence Confirms that Reproductive Risk Factors Can be Used to Stratify Breast Cancer Risks: Implications for a New Population Screening Paradigm." *European Journal of Cancer* 124: 204–06.

Evans, D. Gareth, S. Thomas, J. Caunt, A. Burch, A.R. Brentnall, L. Roberts, A. Howell, M. Wilson, R. Fox, S. Hillier, D.M. Sibbering, S. Moss, M.G. Wallis, D.M. Eccles, FH02 study group, and S. Duffy. 2019. "Final Results of the Prospective FH02 Mammographic Surveillance Study of Women Aged 35-39 at Increased Familial Risk of Breast Cancer." *EClinicalMedicine* 7, 39–46.

Fedewa, Stacey A., Michael Goodman, W. Dana Flanders, Xuesong Han, Robert A. Smith, Elizabeth M. Ward, Chyke A. Doubeni, Ann Goding Sauer, Ahmedin Jemal 2015. "Elimination of Cost-Sharing and Receipt of Screening for Colorectal and Breast Cancer." *Cancer* 121 (18): 3272–80.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker, and the Oregon Health Study Group. 2012. "The Oregon Health Insurance Experiment: Evidence from the First Year." *Quarterly Journal of Economics* 127 (3): 1057–1106.

Habbema, J.D., G.J. van Oortmarssen, D.J. van Putten, J.T. Lubbe, and P. J. van der Maas. 1986. "Age-Specific Reduction in Breast Cancer Mortality by Screening: An Analysis of the Results of the Health Insurance Plan of Greater New York Study." *Journal of the National Cancer Institute* 77 (2): 317–20.

Habermann, Elizabeth B., Beth A. Virnig, Gerald F. Riley, and Nancy N. Baxter. 2007. "The Impact of a Change in Medicare Reimbursement Policy and HEDIS Measures on Stage at Diagnosis among Medicare HMO and Fee-For-Service Female Breast Cancer Patients." *Medical Care* 45 (8): 761–66.

Harding, Charles, Francesco Pompei, Dmitriy Burmistrov, H. Gilbert Welch, Rediet Abebe, and Richard Wilson. 2015. "Breast Cancer Screening, Incidence, and Mortality Across US Counties." *JAMA Internal Medicine* 175 (9): 1483–89.

HCCI (Health Care Cost Institute). 2012. "Health Care Cost and Utilization Report: 2011." https://healthcostinstitute.org/annual-reports/2011-health-care-cost-and-utilization-report.

Hubbard, Rebecca A., Karla Kerlikowske, Chris I. Flowers, Bonnie C. Yankaskas, Weiwei Zhu, and Diana L. Miglioretti. 2011. "Cumulative Probability of False-Positive Recall or Biopsy

Recommendation After 10 Years of Screening Mammography: A Cohort Study." *Annals of Internal Medicine* 155 (8): 481–92.

Humphrey, Linda L., Mark Helfand, Benjamin K.S. Chan, and Steven H. Woolf. 2002. "Breast Cancer Screening: a Summary of the Evidence for the U.S. Preventive Services Task Force." *Annals of Internal Medicine* 137 (5 Part 1): 347–60.

Jacobson, Mireille, and Srikanth Kadiyala. 2017. "When Guidelines Conflict: A Case Study of Mammography Screening Initiation in the 1990s." *Women's Health Issues* 27 (6): 692–99.

Jørgensen, Karsten Juhl, and Peter C. Gøtzsche. 2009. "Overdiagnosis in Publicly Organised Mammography Screening Programmes: Systematic Review of Incidence Trends." *BMJ* 339: b2587.

Jørgensen, Karsten Juhl, Peter C. Gøtzsche, Mette Kalager, and Per-Henrik Zahl. 2017. "Breast Cancer Screening in Denmark." *Annals of Internal Medicine* 167 (7): 524.

Kadiyala, Srikanth, and Erin C Strumpf. 2011. "Are United States and Canadian Cancer Screening Rates Consistent with Guideline Information Regarding the Age of Screening Initiation?" *International Journal for Quality in Health Care* 23 (6): 611–20.

Kadiyala, Srikanth, and Erin C Strumpf. 2016. "How Effective is Population-Based Cancer Screening? Regression Discontinuity Estimates from the US Guideline Screening Initiation Ages." *Forum for Health Economics & Policy* 19 (1): 87–139.

Kelaher, M., and J.M. Stellman. 2000. "The Impact of Medicare Funding on the Use of Mammography among Older Women: Implications for Improving Access to Screening." *Preventive Medicine* 31 (6): 658–64.

Kim, Hyuncheol Bryant, and Sun-Mi Lee. 2017. "When Public Health Intervention is not Successful: Cost Sharing, Crowd-Out, and Selection in Korea's National Cancer Screening Program." *Journal of Health Economics* 53: 100–16.

Kolata, Gina. 2009. "Get a Mammogram. No Don't. Repeat." *New York Times*. https://www.nytimes.com/2009/11/22/weekinreview/22kolata.html

Kowalski, Amanda E. 2019: Behavior within a Clinical Trial and Implications for Mammography Guidelines. Working Paper No. 25049, National Bureau of Economic Research.

Lee, Nancy C., Faye L. Wong, Patricia M. Jamison, Sandra F. Jones, Louise Galaska, Kevin T. Brady, Barbara Wethers, and George-Ann Stokes-Townsend. 2014. "Implementation of the National Breast and Cervical Cancer Early Detection Program: The beginning." *Cancer* 120 (S16): 2540–48.

Levano, Whitney, Jacqueline W. Miller, Banning Leonard, Linda Bellick, Barbara E. Crane, Stephenie K. Kennedy, Natalie M. Haslage, Whitney Hammond, and Felicia S. Tharpe. 2014. "Public Education and Targeted Outreach to Underserved Women Through the National Breast and Cervical Cancer Early Detection Program." *Cancer* 120 (S16): 2591–96.

Lu, Yao, and David J.G. Slusky. 2016. "The Impact of Women's Health Clinic Closures on Preventive Care." *American Economic Journal: Applied Economics* 8 (3): 100–24.

Maciosek, Michael V., Ashley B. Coffield, Thomas J. Flottemesch, Nichol M. Edwards, and Leif I. Solberg. 2010. "Greater Use of Preventive Services in U.S. Health Care Could Save Lives at Little or No Cost." *Health Affairs* 29 (9): 1656–60.

Marmot, M.G., D.G. Altman, D.A. Cameron, J.A. Dewar, S.G. Thompson, M. Wilcox, and The Independent UK Panel on Breast Cancer Screening. 2013. "The Benefits and Harms of Breast Cancer Screening: An Independent Review." *British Journal of Cancer* 108 (11): 2205–40.

Mehta, Shivan J., Daniel Polsky, Jingsan Zhu, James D. Lewis, Jonathan T. Kolstad, George Loewenstein, and Kevin G. Volpp. 2015. "ACA Mandated Elimination of Cost Sharing for Preventive Screening Has Had Limited Early Impact." *American Journal of Managed Care* 21 (7): 511–17.

Miller, Anthony B., Teresa To, Cornelia J. Baines, Claus Wall. 2000. "The Canadian National Breast Screening Study-2: 13-Year Results of a Randomized Trial in Women Aged 50–59 Years." *Journal of the National Cancer Institute* 92 (18): 1490–99.

Miller, Anthony B., Teresa To, Cornelia J. Baines, Claus Wall. 2002. "The Canadian National Breast Screening Study-1: Breast Cancer Mortality after 11 to 16 Years of Follow-up: A Randomized Screening Trial of Mammography in Women Age 40 to 49 Years." *Annals of Internal Medicine* 137 (5 Part 1): 305.

Moss, Sue M., Howard Cuckle, Andy Evans, Louise Johns, Michael Waller, Lynda Bobrow, and Trial Management Group. 2006. "Effect of Mammographic Screening from Age 40 Years on Breast Cancer Mortality at 10 Years' Follow-Up: A Randomised Controlled Trial." *Lancet* 368 (9552): 2053–60.

Mukama, Trasias, Mahdi Fallahad, Yu Tian, Kristina Sundquist, Jan Sundquist, Hermann Brenner, and Elham Kharazmi. 2020. "Risk-Tailored Starting Age of Breast Cancer Screening Based on Women's Reproductive Profile: A Nationwide Cohort Study." *European Journal of Cancer* 124: 207–13.

Nelson, Heidi D., Kari Tyne, Arpana Naik, Christina Bougatsos, Benjamin Chan, Peggy Nygren, and Linda Humphrey. 2009. "Screening for Breast Cancer: Systematic Evidence Review Update for the U. S. Preventive Services Task Force." *Annals of Internal Medicine* 151 (10): 727–W242.

Nyström, Lennarth, Ingvar Andersson, Nils Bjurstam, Jan Frisell, Bo Nordenskjöld, and Lars Erik Rutqvist. 2002. "Long-Term Effects of Mammography Screening: Updated Overview of the Swedish Randomised Trials." *Lancet* 359 (9310): 909–19.

Oeffinger, Kevin C., Elizabeth T. H. Fontham, Ruth Etzioni, Abbe Herzig, James S. Michaelson, Ya-Chen Tina Shih, Louise C. Walter, Timothy R. Church, Christopher R. Flowers, Samuel J. LaMonte, Andrew M. D. Wolf, Carol DeSantis, Joannie Lortet-Tieulent, Kimberly Andrews, Deana Manassaram-Baptiste, Debbie Saslow, Robert A. Smith, Otis W. Brawley, and Richard Wender. 2015. "Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update from the American Cancer Society." *Journal of the American Medical Association* 314 (15): 1599–1614.

Ong, Mei-Sing, and Kenneth D. Mandl. 2015. "National Expenditure for False-Positive Mammograms and Breast Cancer Overdiagnoses Estimated at $4 Billion a Year." *Health Affairs* 34 (4): 576–83.

Oster, Emily. 2020. "Health Recommendations and Selection in Health Behaviors." *American Economic Review: Insights* 2 (2): 143-60.

O'Sullivan, Jennifer, Celinda Franco, Beth C. Fuchs, Bob Lyke, Richard Price, and Kathleen S. Swendiman. 1997. "Medicare Provisions in the Balanced Budget Act of 1997." CRS Report for Congress 97-802. https://www.everycrsreport.com/files/19970818_97-802_0347a1fe 70af1b6bbce038af2c217186942cd7cc.pdf.

Persico, Nicola. 2009: "Racial Profiling? Detecting Bias Using Statistical Evidence." *Annual Review of Economics* 1 (1): 229–54.

Saad, Lydia. 2009. "Women Disagree with New Mammogram Advice." *Gallup.* http://news. gallup.com/poll/124463/wo disagree-new-mammogram-advice.aspx.

Schousboe, John T., Karla Kerlikowske, Andrew Loh, and Steven R. Cummings. 2011. "Personalizing Mammography by Breast Density and Other Risk Factors for Breast Cancer: Analysis of Health Benefits and Cost-Effectiveness." *Annals of Internal Medicine* 155 (1): 10-20.

SEER (Surveillance, Epidemiology, and End Results Program). 2019. "SEER Incidence Data, 1973-2015." https://seer.cancer.gov/data/index.html.

Segel, Joel E., Rajesh Balkrishnan, and Richard A. Hirth. 2017. "The Effect of False-Positive Mammograms on Antidepressant and Anxiolytic Initiation." *Medical Care* 55 (8): 752–58.

Susan G. Komen Foundation. 2018: "Accuracy of Mammograms." https://ww5.komen.org/ BreastCancer/Accuracyofl

Tan, Sita Y.G.L., Gerrit J. van Oortmarssen, Harry J. de Koning, Rob Boer, J. Dik F. Habbema. 2006. "The MISCAN-Fadia Continuous Tumor Growth Model for Breast Cancer.". *Journal of the National Cancer Institute, Monographs* 36: 56–65.

Taubes, Gary..1997. "The Breast-Screening Brawl." *Science* 275 (5303): 1056–59.

Vang, Suzanne, Laurie R. Margolies, and Lina Jandorf. 2018. "Mobile Mammography Participation Among Medically Underserved Women: A Systematic Review." *Preventing Chronic Disease* 15: E140.

Welch, H. Gilbert. 2015. *Less Medicine, More Health: 7 Assumptions That Drive Too Much Medical Care*. Beacon Press.

Welch, H. Gilbert, and William C. Black. 2010. "Overdiagnosis in Cancer." *Journal of the National Cancer Institute* 102 (9): 605–13.

Welch, H. Gilbert, and Honor J. Passow. 2014. "Quantifying the Benefits and Harms of Screening Mammography." *JAMA Internal Medicine* 174 (3): 448–54.

Welch, H. Gilbert , Philip C. Prorok, A. James O'Malley, and Barnett S. Kramer. 2016. "Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness.".*New England Journal of Medicine* 375 (15): 1438–47.

Welch, H. Gilbert, Lisa M. Schwartz, and Steven Woloshin. 2011. *Overdiagnosed: Making People Sick in the Pursuit of Health*. Beacon Press.

Zackrisson, Sophia, Ingvar Andersson, Lars Janzon, Jonas Manjer, and Jens Peter Garne. 2006. "Rate of Over-Diagnosis of Breast Cancer 15 Years After End of Malmö Mammographic Screening Trial: Follow-Up Study." *BMJ* 332 (7543): 689–92.

Zahl, Per-Henrik, Jan Maehlen, and H. Gilbert Welch. 2008. "The Natural History of Invasive Breast Cancers Detected by Screening Mammography." *Archives of Internal Medicine* 168 (21), 2311–16.

Figure 1: Mammogram rates by age

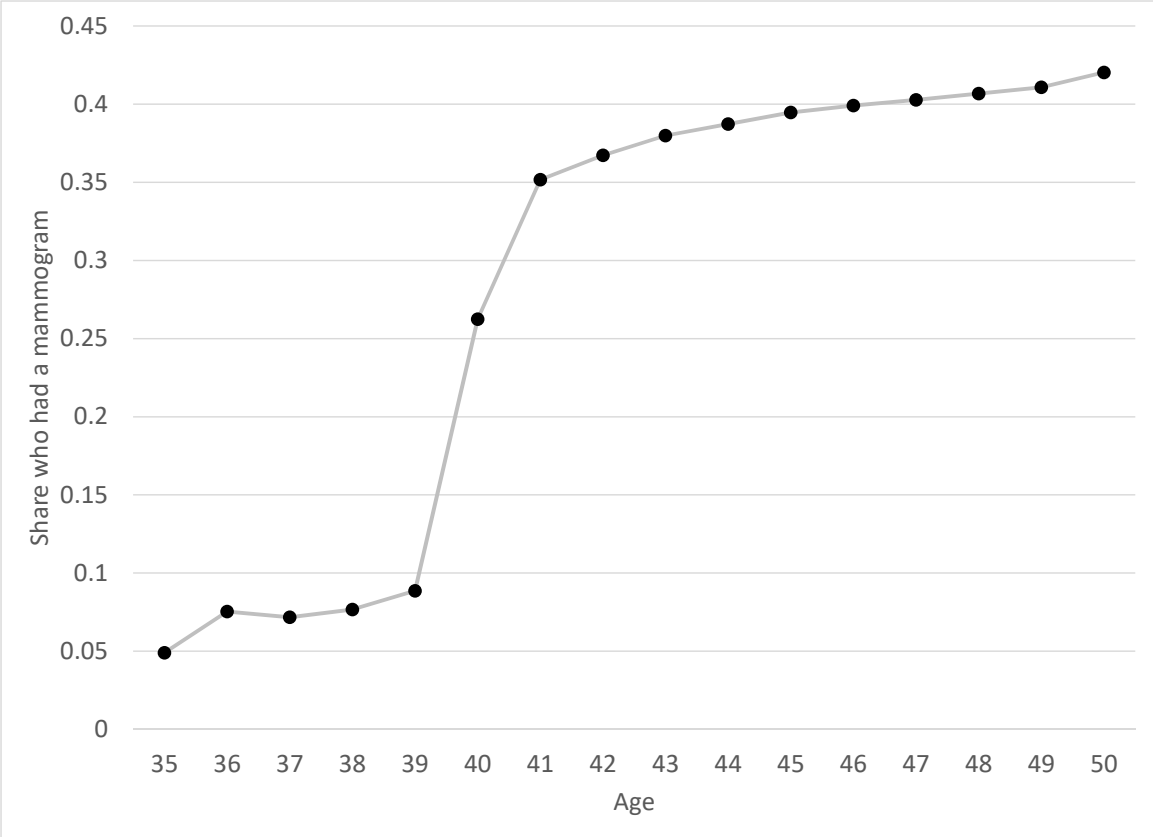
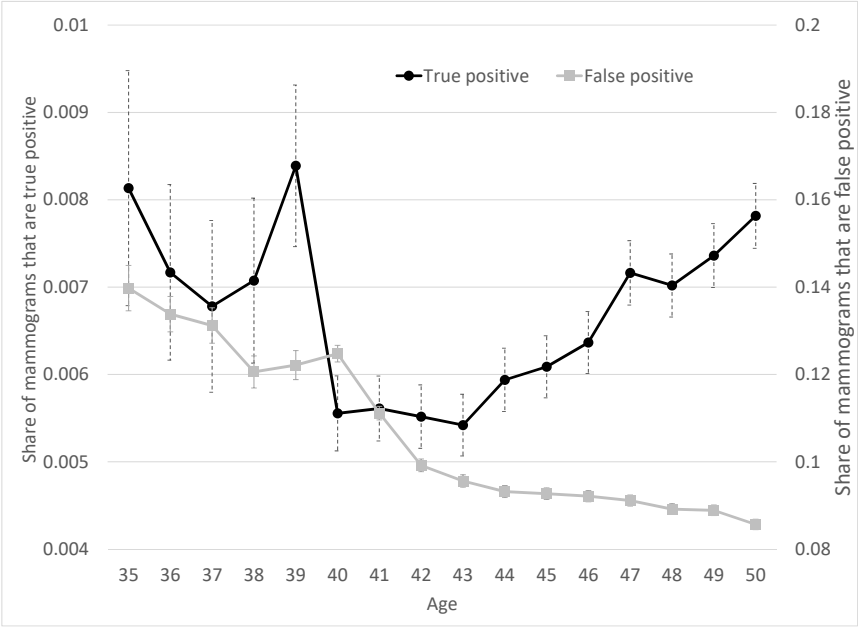


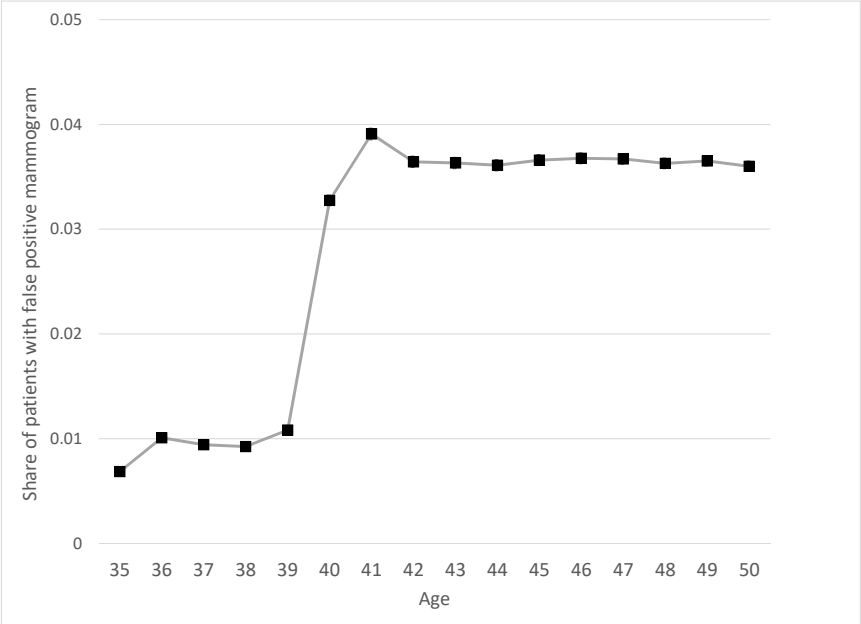
Figure shows share of women who had a mammogram by age, from insurance claims data on a set of privately insured woman-years from 2008-2012, for mammograms between 2009-2011. Because we observe birth year, age is measured as of the start of the calendar year. Thus the mammogram rate at age 40 is the share of women who got a mammogram in the year they turned 40. Error bars (small, and therefore not visible in the figure) reflect 95% confidence intervals. N = 7,373,302 woman-years.

Figure 2: Mammogram outcomes by age

(a) Share of mammograms that are true positive and false positive



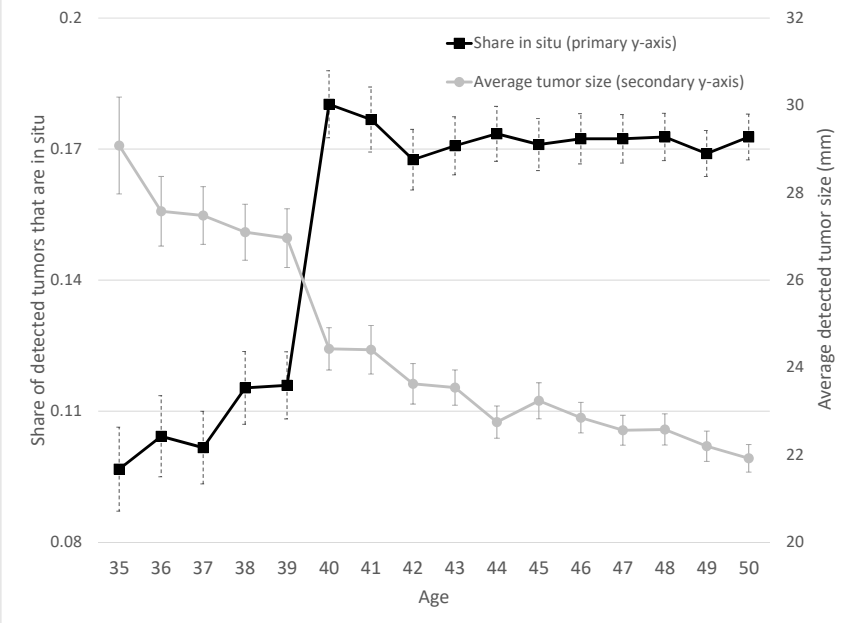
(b) Share of women with a false positive by age



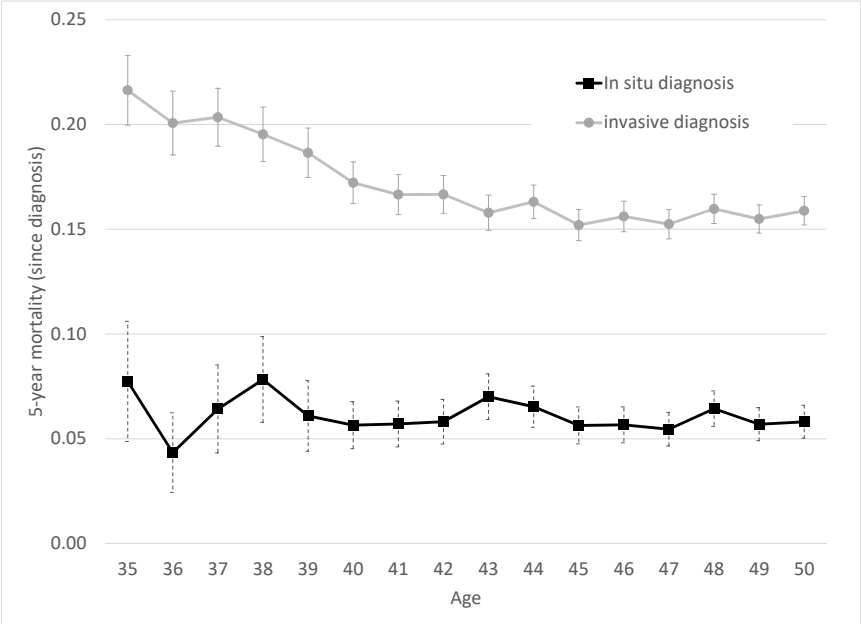
Sample is limited to the set of privately insured woman-years from the private insurance claims data who had a mammogram. N = 7,373,302 woman-years. For each age (measured by the age at the beginning of the calendar year), panel (a) shows the share of mammograms that are true positive (left hand axis) and false positive (right hand axis); the omitted category is mammograms that are negative. Panel (b) presents the share of women with a false positive by age; this reflects both mammogram rates by age from Figure 1, and the share of mammograms with a false positive by age from panel (a). Error bars reflect 95% confidence intervals.

Figure 3: Tumor characteristics and mortality by age

(a) Tumor stage and size by age



(b) Mortality



Panel (a) shows diagnosed breast cancer tumors by age in the SEER data from 2000-2015; N =197,956 breast cancer diagnoses. Primary y-axis shows share of breast cancer tumors that are in-situ; secondary y-axis shows average size of diagnosed tumors. Panel (b) shows 5-year mortality for diagnosed breast cancer tumors separately by age of diagnoses and by tumor stage (in-situ and invasive) in the SEER data from 2000-2010 to account for five-year mortality outcomes by 2015; N = 147,243 diagnoses with non-missing 5-year mortality. Error bars reflect 95% confidence intervals.

Figure 4: Characteristics of who selects into mammograms

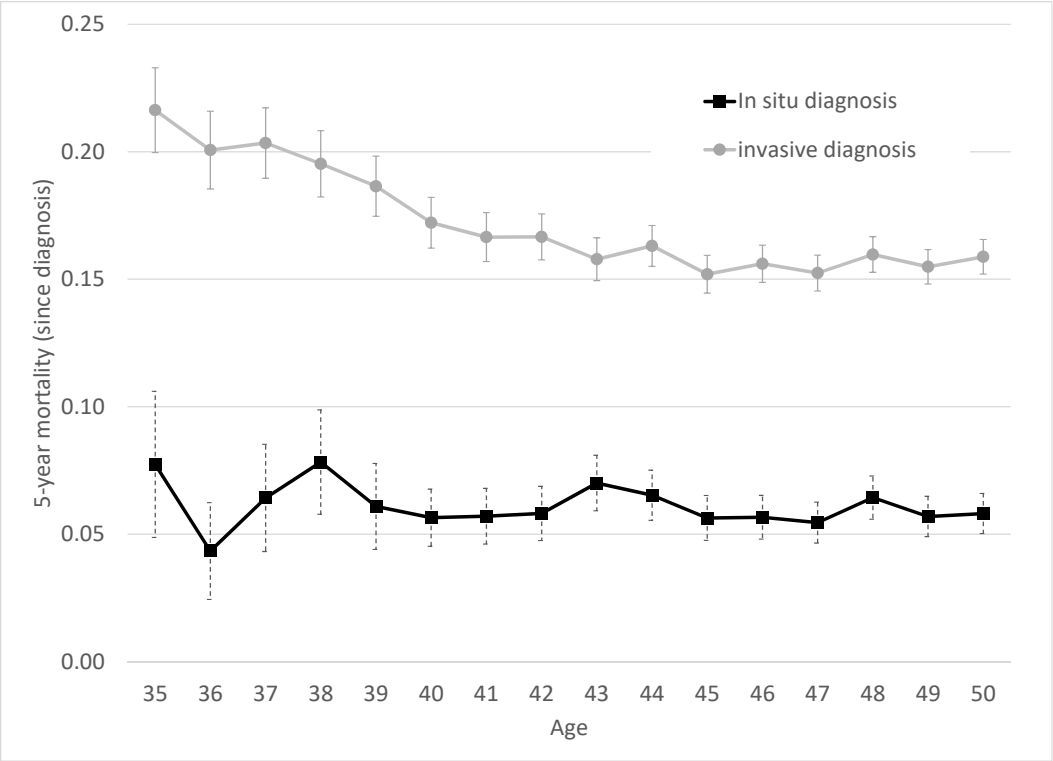
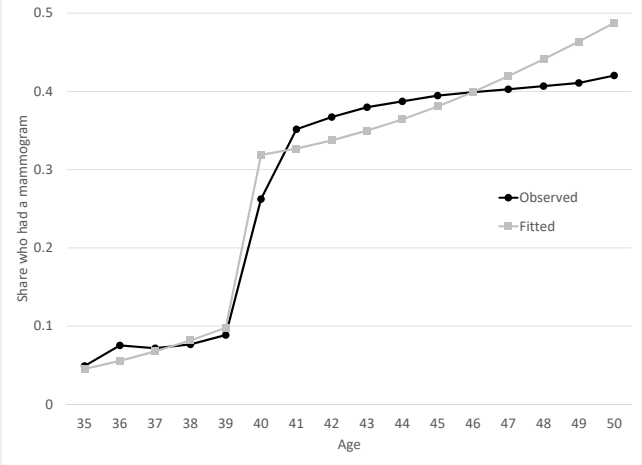


Figure reports the ratio of health care use, behavior, and demographics for compliers relative to always-takers (left panel) and compliers relative to never-takers (right panel). The mean characteristics for these groups were calculated using regression coefficients from the estimation of equation (A.1) as described in Appendix B. Error bars represent 95% confidence intervals. Standard errors are constructed using a bootstrap with 100 repetitions clustered at the age level. The error bars for "Use Oral Contraceptive Pill" and "# Drinks in Prior Month" in the right panel are truncated at zero and two for scaling; the actual bootstrap confidence intervals are larger. The sample in the first section is a set of privately insured woman-years from HCCI from 2008-2012, for mammograms between 2009-2011. The sample in the second and third sections is from BRFSS for even years 2000-2012, restricted to women with any health insurance (the data do not distinguish between public or private insurance status). Details for each outcome are listed in Appendix Figures A.3, A.4, and A.5.

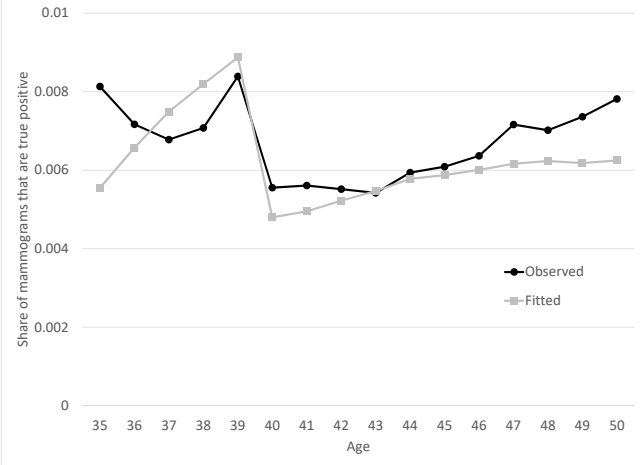
Figure reports the ratio of health care use, behavior, and demographics for compliers relative to always-takers (left panel) and compliers relative to never-takers (right panel). The mean characteristics for these groups were calculated using regression coefficients from the estimation of equation (A.1) as described in Appendix B. Error bars represent 95% confidence intervals. Standard errors are constructed using a bootstrap with 100 repetitions clustered at the age level. The error bars for "Use Oral Contraceptive Pill" and "# Drinks in Prior Month" in the right panel are truncated at zero and two for scaling; the actual bootstrap confidence intervals are larger. The sample in the first section is a set of privately insured woman-years from HCCI from 2008-2012, for mammograms between 2009-2011. The sample in the second and third sections is from BRFSS for even years 2000-2012, restricted to women with any health insurance (the data do not distinguish between public or private insurance status). Details for each outcome are listed in Appendix Figures A.3, A.4, and A.5.

Figure 5: Model fit

(a) Share with mammogram



(b) Share of mammograms that are true positive



(c) Share of diagnoses that are in-situ

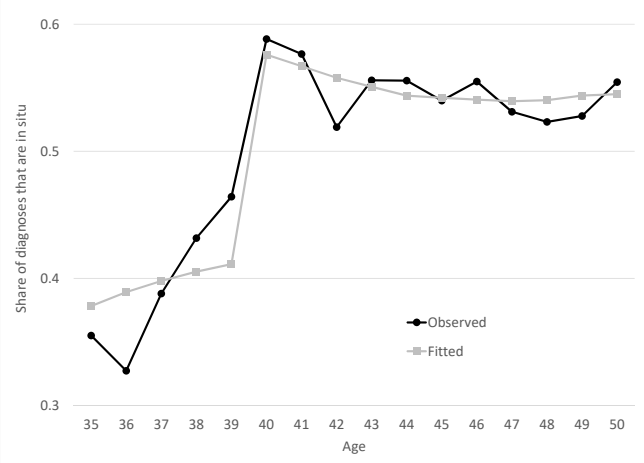
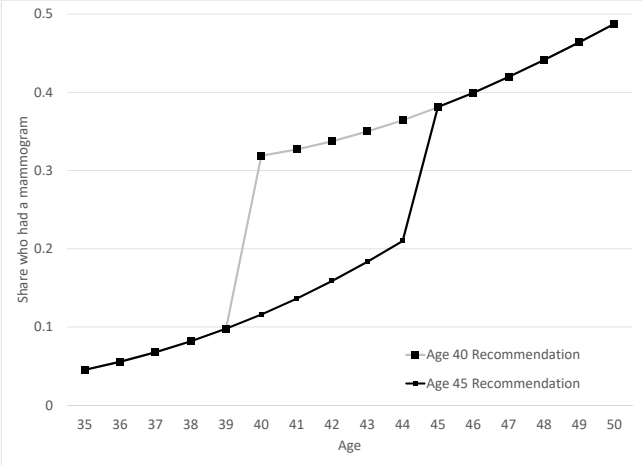


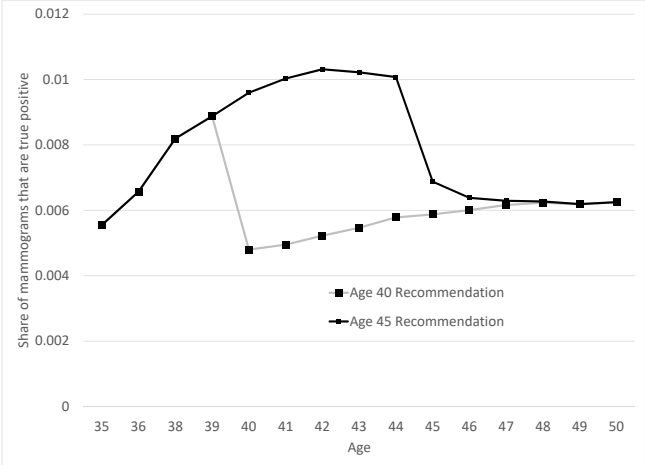
Figure shows model fit by comparing the observed patterns of mammogram rates, outcomes, and types of diagnoses by age to the fitted values from the model based on the parameter estimates from Table 2. The observed data on mammograms (Panel (a)) was previously shown in Figure 1; the observed data on share of mammograms that are true positives was previously shown in Figure 2a; the observed data on the share of diagnoses that are in-situ is a modified version of the data shown in Figure 3a. While Figure 3a presented the share of all diagnosed cancers that are in-situ, we match the share of mammogram-diagnosed cancers that are in-situ, as shown in Panel (c). Appendix D provides more detail.

Figure 6: Impact of changing the mammogram recommendation age from 40 to 45, by age

(a) Share with mammogram



(b) Share of mammograms that are true positive



(c) Share of diagnoses that are in-situ

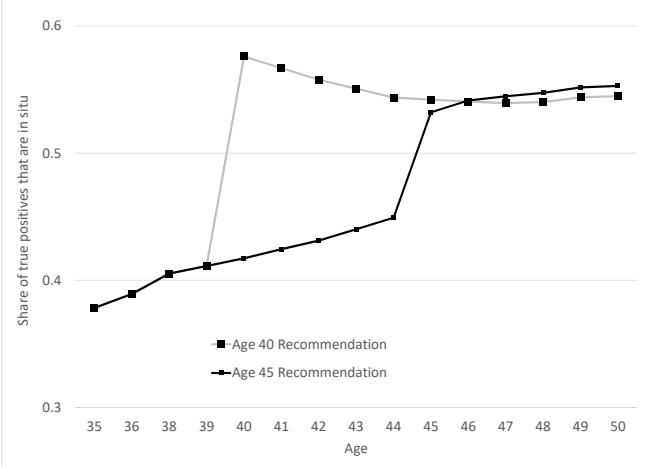


Figure reports the model predictions - by age - for mammogram rates, mammogram outcomes, and the share of diagnoses that are in-situ, based on the parameter estimates from Table 2. As in Table 3, we report the model predictions both under the status quo recommendation that mammograms begin at age 40 and the counterfactual recommendation that mammograms begin at age 45.

Table 1: Summary statistics

	No. of Observations		Health Care Spending	
	N (000s)	Share	Total	Out-of-pocket
No mammogram	5,166.2	0.701	$4,300	$625
Mammogram	2,206.9	0.299	$4,985	$751
Conditional on mammogram:				
Negative	1,977.8	0.896	$4,552	$715
False positive	214.6	0.097	$6,106	$952
True Positive	14.4	0.007	$47,639	$2,821

Table shows summary statistics from insurance claims data on a set of 35-50 year old privately insured women from 2008-2012, for mammograms between 2009-2011. Each observation is a woman-year. 12-month spending measures health care spending in the 12 months after the mammogram (including the mammogram itself) for those with a mammogram. For those without a mammogram, we draw a reference date from the distribution of actual mammograms in that year. All reference dates are set to be the first of the given month. Spending is measured in the 12 months after this reference date.

Table 2: Parameter estimates

Parameter	Estimate	95% Confidence Interval
α^o	-5.21	[-5.63 , -4.48]
γ^o	0.10	[0.08 , 0.11]
$\delta^o_{\text{in-situ}}$	0.36	[0.29 , 0.97]
$\delta^o_{\text{invasive}}$	1.13	[0.98 , 56.73]
α^r	0.29	[-0.63 , 1.18]
γ^r	-0.03	[-0.05 , 0.00]
$\delta^r_{\text{in-situ}}$	-0.01	[-0.20 , 0.77]
$\delta^r_{\text{invasive}}$	-4.67	[-142.97 , -0.01]

Table shows the parameter estimates from the mammogram decision model. Confidence intervals are calculated using 100 repetitions of the bootstrap.

Table 3: Impact of changing the mammogram recommendation age from 40 to 45

	Rec at Age 40	Rec at Age 45	Change
A. Screening and spending (per woman)			
Mammograms	4.70	3.80	-0.90
	(0.06)	(0.14)	(0.08)
Negative	4.22	3.42	-0.81
	(0.05)	(0.12)	(0.07)
False positives	0.46	0.36	-0.09
	(0.01)	(0.02)	(0.01)
True positives	0.0208	0.0204	-0.0004
	(0.0024)	(0.0024)	(0.0001)
In-situ diagnoses	0.0063	0.0060	-0.0004
	(0.0005)	(0.0005)	(0.0001)
Invasive diagnoses	0.0145	0.0145	0.0000
	(0.0019)	(0.0019)	(0.0001)
Total health care spending ($)	71,326	71,007	-319
	(128)	(155)	(29)
B. Mortality (per 1,000 women by age 50)			
Dead	15.98	16.03	0.05
	(0.53)	(0.53)	(0.03)
Dead from breast cancer	8.23	8.28	0.05
	(0.53)	(0.53)	(0.03)
Dead from other reason	7.75	7.75	0.00
	(0.00)	(0.00)	(0.00)
Years alive, per woman	15.87	15.87	-0.0002
	(0.00)	(0.00)	(0.0001)

Table reports model predictions for various outcomes under the status quo recommendation that mammograms begin at age 40 (column 1) and the counterfactual recommendation that mammograms begin at age 45 (column 2). The predictions are generated using the parameter estimates from Table 2, and simulated women's life histories under a non-screening regime based on the clinical oncology model. Panel A reports the average number of mammograms and different mammogram outcomes per woman over ages 35-50. Panel B shows the share of women dead (and from different causes) by age 50, as well as the number of years alive on average between 35 and 50. Standard errors are calculated using 100 repetitions of the bootstrap.

Table 4: Spending differences for different components of spending

	Recommendation at Age 40	Recommendation at Age 45	Difference
A. Estimated Selection			
Mammograms (per woman)	4.70	3.80	-0.90
	(0.06)	(0.14)	(0.08)
Total health care spending ($ per woman)	71,326	71,007	-319
	(128)	(155)	(29)
Dead by age 50 (per 1,000 women)	15.98	16.03	0.05
	(0.53)	(0.53)	(0.03)
B. No Selection			
Mammograms (per woman)	4.70	3.80	-0.90
	(0.06)	(0.14)	(0.08)
Total health care spending ($ per woman)	71,364	71,024	-340
	(111)	(147)	(37)
Dead by age 50 (per 1,000 women)	15.84	16.02	0.18
	(0.47)	(0.53)	(0.06)
C. Consistent Selection			
Mammograms (per woman)	4.70	3.80	-0.90
	(0.06)	(0.14)	(0.08)
Total health care spending ($ per woman)	71,450	71,068	-382
	(87)	(134)	(48)
Dead by age 50 (per 1,000 women)	15.54	15.99	0.45
	(0.39)	(0.52)	(0.13)

Table reports model predictions under the status quo recommendation that mammograms begin at age 40 (column 1) and the counterfactual recommendation that mammograms begin at age 45 (column 2). Each panel reports results under different assumptions about the nature of selection both in the absence and presence of a recommendation. Panel A reports results based on the estimated selection patterns; these results repeat findings shown previously in Table 3. Panel B repeats the same exercises as in Panel A, but instead of using the estimated selection (i.e. the δ^o and δ^r parameters shown in Table 2), we instead assume "no selection" (i.e. we set $\delta^o = \delta^r = 0$). Panel C also repeats the exercises in Panel A but now assumes "consistent selection" (i.e. we set δ^r equal to our estimates of δ^o in Table 2). In both Panel B and C, we hold the overall mammogram rate fixed at Panel A's predicted age-specific mammogram rates (which of course varies in column 1 and column 2), so that the counterfactuals across panels consider differences in selection, not in levels. To do this we adjust the intercept α_r for each age and counterfactual to match the age-specific mammogram rates in Panel A, assuming the simulated life histories and cancer status remains constant. The small differences in mammograms in Panel A and Panel C are due to changes in the denominator of simulated life histories. Specifically, since fewer women die in Panel C, there are more years where they could potentially obtain a mammogram. Standard errors are calculated using 100 repetitions of the bootstrap.

Chapter 2

The Effects of Floodplain Regulation on Housing Markets

By Abigail Ostriker and Anna Russo[1]

Abstract

We investigate the effects of housing regulations designed to correct a wedge between privately- and socially-optimal construction in areas at risk of flooding in Florida. Using a spatial regression discontinuity around regulatory boundaries and an event study around the policy's introduction, we document that floodplain regulation reduces new construction in high-risk areas and increases the share of newly-built houses that are elevated. Embedding these effects in a model of residential choices with elastic housing supply, we find that the policy reduces expected flood damages by 60%. One-quarter of this reduction is driven by relocation of new construction to lower-risk areas, and three-quarters is driven by elevation of houses remaining in risky areas. However, this second-best policy achieves at best about one-tenth of possible welfare gains because of poor targeting. It overcorrects in many areas, inducing more consumers to elevate and relocate than is socially-optimal, while still allowing inefficiently-high construction in the riskiest places. By contrast, a flexible corrective tax on flood risk would achieve substantial welfare gains of more than $2,700 per newly-developed house.

2.1 Introduction

Floods cause an estimated $32 billion in damages per year, making them the costliest form of natural disaster in the United States (Wing et al., 2022). These losses are expected to grow by 25% by 2050, reflecting both an increase in flood hazard due to climate change and a concentration of population growth in risky locations

[1]Ostriker: Department of Economics, MIT, ostriker@mit.edu; Russo: Department of Economics, MIT, aerusso@mit.edu.

We are grateful to Nikhil Agarwal, Clare Balboni, Amy Finkelstein, and Ben Olken for invaluable guidance, mentorship, and advice. We also thank Parinitha Sastry, Patrick Schwarz, Katherine Wagner, and participants in the MIT PF lunch, IO lunch and Energy/Environmental tea for helpful comments. We benefited from the advice of Jennie Murack, Daniel Sheehan, and Jeff Blossom regarding GIS and the process of digitizing maps. We thank Cathy Yung, Jerry Li, and the associates at DIN Engineering for helpful research assistance. We also thank Milani Chatterji-Len, Luke Fredrickson, Robert Wiley, and Howard Critchfield for providing useful context about flood maps and floodplain regulations, and First Street Foundation for sharing their estimates of flood risk with us. We gratefully acknowledge support from the C. Lowell Harriss Dissertation Fellowship and from the MIT Economics Department's George and Obie Shultz Fund. This material is based upon work supported by the National Science Foundation Graduate Fellowship Program under grant number 1745302.

(Wing et al., 2022). A central concern among policymakers and economists is that these trends are driven by a wedge between social and private values for flood safety (Kydland and Prescott, 1977; Coate, 1995; Ben-Shahar and Logue, 2016). Indeed, a large body of work has documented that private incentives to reduce flood risk are muted by both misperceptions of that risk and expectations of government aid (Gallagher, 2014; Kousky et al., 2018b; Bakkensen and Barrage, 2021; Wagner, 2021; Mulder, 2021). To correct these market frictions, policymakers demarcate especially risky locations as "Special Flood Hazard Areas" (SFHAs) and regulate them more strictly. Inside the SFHA, developers are required to build elevated homes, and homeowners face a flood insurance purchase mandate and high flood insurance prices.[2]

This paper investigates the impact of floodplain regulation on the location of new construction, housing prices, estimated flood damages, and social welfare. The effect of this coarse policy on welfare is ambiguous, as it might not reduce damages more than the costs it imposes via distortions in the housing market. In this paper, we study the extent to which floodplain regulation reduces flood damages via both the location and elevation of construction, and we weigh these benefits against the costs of that regulation. We do so by studying effects around the boundaries of regulation and around the date of the regulation's introduction. We then embed these empirical results in a model of residential choices and development to extrapolate effects away from the boundary, estimate the costs of regulation, and investigate the welfare effects of current and counterfactual policies.

To conduct our analysis, we assemble a new comprehensive and spatially-granular dataset to describe regulation, real estate development, and flood risk in Florida over a 40-year time horizon. Florida is both populous and highly flood-prone, with 45% of land currently designated as a high-risk flood zone. Our dataset combines maps of historic flood zone extents, granular remote-sensing-based measures of historic and current development, administrative data describing house prices and attributes, and the flood risk profile of current and counterfactual development, generated from a state-of-the-art hydrological model. The scope and detail of our dataset are both critical to our analysis. Because our dataset extends to the first maps delineating regulated areas — which we digitized from archival scans for this project — we are able to study the policy's effect on long-run development and confirm the validity of our core empirical strategy. And because our dataset details both elevation and location decisions, as well as how flood damages vary along these margins, we are able to comprehensively measure the policy's effect on flood risk.

We use two complementary empirical strategies — a spatial regression discontinuity and an event study around the regulation's implementation — to characterize the policy's risk reduction effects along two margins: reduced construction and mandatory elevation in risky areas. Our spatial regression discontinuity design compares current development and house prices on either side of the regulatory boundary delineated at the time of the policy's introduction in the 1970s and 1980s. Our analysis relies on the assumption that flood risk and other amenities are smooth through these initial regulatory boundaries. While unlikely to hold in modern maps, this assumption is reasonable for the original maps because mapping technologies were rudimentary and homeowners lacked the ability to influence the initial regulatory boundaries.[3] Importantly, we validate this assumption by examining smoothness in pre-period land use through the historic boundaries. We document that modern-day development is 9% lower just inside the regulated area, highlighting the

[2]The policy instrument of creating a binary distinction of "floodplain" or not and imposing both insurance and building requirements is not unique to the United States. EU countries and Australia also manage flood risk via the creation of flood maps that influence both flood insurance and building codes (de Moel et al., 2009; Golnaraghi et al., 2020).

[3]After the initial maps were drawn, landowners could deregulate developed parts of their properties either by petitioning to correct a mistake or physically elevating land to reduce its risk. Since maps are updated over time, this behavior produces a negative correlation between development and floodplain designation in modern flood maps.

potential for the policy to reduce damages by shifting development out of risky areas. This decrease in development is not accompanied by a reduction in prices – if anything, prices are slightly higher just inside the regulated area – indicating that floodplain regulation imposes costs on developers, which they at least partially pass through to consumers.

We also document reduced damages on the intensive margin, through elevation of houses that are built in the floodplain. Building on Wagner (2021), we exploit the sharp timing of the policy's introduction to examine the effect of building standards on flood risk in an event study design. We find that regulation increases the share of elevated homes by 27 percentage points and reduces flood risk by 3% of average home value. Though mandatory elevation generates social value via reduced damages, we show that homeowners do not privately value this reduction in flood risk. This result is consistent with prior work documenting a large wedge between social and private valuations of flood risk, implying scope for welfare-improving intervention. It also suggests that the increase in price across the boundary is attributable to construction costs, not differences in willingness-to-pay for elevated homes.

Together, our empirical results yield four facts. First, floodplain regulation suppresses construction in high-risk areas. Second, the policy's elevation mandate binds, reducing flood risk. Third, the cost of mandating elevation is large enough that the price effect of the inward supply shift dominates at the boundary. Fourth, our setting exhibits a wedge between private and social valuations for flood safety, yielding the potential for the policy to improve welfare. However, these results alone are insufficient to quantify the total effect of the policy on either damages or overall welfare. The policy's effects depend critically on the location of counterfactual development: both risk and amenities vary across space. Additionally, the large size of the regulated area may affect house prices in the unregulated locations, which could mitigate the incentive to relocate and impose pecuniary externalities on unregulated consumers. Moreover, a full welfare analysis of the policy requires us to quantify the regulation's costs to developers and consumers.

We therefore specify and estimate a model of residential choice and real estate development. In our model, individuals maximize utility when choosing census-tract-by-flood-zone locations, as a function of prices, floodplain status, and unobserved amenities. Landowners build houses when doing so is more profitable than their outside option of land use; housing profits depend on housing prices, tract-zone-specific construction costs, and costs of elevation. Because our event study estimates imply that homeowners are unwilling to pay for elevation, we simplify the model by differentiating houses only by location — not risk — and assuming that suppliers do not endogenously provide elevated houses absent policy intervention. We estimate the model's key parameters — demand and supply costs of floodplain designation — by matching the cross-border differences we find in our boundary discontinuity analysis. We calibrate supply and demand price elasticities to estimates from the literature (Calder-Wang, 2021; Song, 2021; Baum-Snow and Han, 2022).

We first use the model to quantify the policy's impact on expected flood damages. We find that the policy reduces expected flood damages by 60%, or approximately $3.5 billion per county. Both the extensive-margin location channel and intensive-margin elevation channel are quantitatively important in this setting. We estimate that one-quarter of the reduction in damages is driven by the relocation of new construction to lower-risk locations. The remaining three-quarters are due to elevation mandates for the houses that remain in risky locations. However, we find that these risk reductions are driven by large costs imposed by the regulation on both consumers and producers of housing. Our parameter estimates suggest that mandatory elevation of a house increases construction costs by 25%. Furthermore, consumers are willing to pay 23% more for an equivalent house to avoid living in a regulated area.

We conclude by developing a normative framework that allows us to estimate and compare welfare under current and counterfactual policies. We define social welfare to capture both producer and consumer surplus and any socially-costly damages that consumers do not value, whether due to internalities or externalities. Because the wedge between private and social valuation of flood risk could be due to misperceptions, our welfare framework allows consumers' decision and experienced utility to diverge (Allcott and Taubinsky, 2015). We contrast the status quo policy regime against both an unregulated benchmark and a first-best corrective tax in the spirit of Pigou (1920) equal to the social cost of flooding, which varies by location and elevation status. This corrective tax is not only a useful theoretical benchmark, but a plausible extension of recent policy changes which have introduced more property-specific granularity in flood insurance premiums (Flavelle, 2021).

Though the current policy achieves approximately the socially-efficient degree of damage reductions, we estimate that it achieves at best about 10% of the possible social welfare gains from regulation. The policy performs poorly despite correcting market frictions because the uniform delineation of high-risk zones is poorly-targeted. In some regulated but relatively safe areas, the current policy is over-corrective: the elevation and relocation it induces is costlier than the benefits it generates, and this adaptation would not occur under the first-best regulation. Meanwhile, in the riskiest areas, the current policy should be even stronger: expected damages are so large that reducing risk is worth more than even the large costs imposed by current regulation. These findings underscore the gains to transitioning from a blunt attribute-based regulation to a more flexible policy instrument, such as a tax or an actuarially-fair insurance mandate, that can efficiently incentivize both elevation and relocation where appropriate.

Our work builds on and contributes to a number of distinct literatures. Most narrowly, we contribute to a body of work analyzing various aspects of floodplain regulation, including its effect on house prices (Harrison et al., 2001; Bin and Landry, 2013; Indaco et al., 2018; Gibson and Mullins, 2020; Hino and Burke, 2021; Lee, 2022), in-place adaptation (Wagner, 2021; Mulder, 2021), and population (Lee, 2022; Peralta and Scott, 2019). We build on this literature by developing an equilibrium framework that allows us to simultaneously analyze and quantify both the extensive and intensive channels for averting damages, and, importantly, to trade them off against regulatory costs.

We also contribute to a literature documenting frictions in mitigating or adapting to climate risk (Annan and Schlenker, 2015; Deryugina and Kirwan, 2018; Kousky et al., 2018b; Balboni, 2019; Baylis and Boomhower, 2019; Bakkensen and Barrage, 2021; Wagner, 2021; Mulder, 2021). In particular, we investigate the extent to which a second-best corrective policy can reduce the welfare losses from these frictions. In doing so, our paper relates to a body of work examining second-best corrective instruments to reduce welfare losses from externalities or internalities (Ito and Sallee, 2018; Germeshausen, 2018; Barahona et al., 2020; Kellogg, 2020).

Methodologically, we draw on work in urban economics that embeds boundary discontinuity designs in discrete choice frameworks (Bayer et al., 2007; Anagol et al., 2021; Song, 2021). We build on these models by incorporating recent estimates of spatially-granular supply elasticities from Baum-Snow and Han (2022) to more-accurately characterize equilibrium changes in housing supply across markets.

The next section describes the institutional details of the National Flood Insurance Program, including the regulations imposed inside the SFHA and the process of generating flood maps that distinguish between SFHA and non-SFHA land. The following section describes our setting — the state of Florida — and data. In Section 2.4, we present reduced-form evidence of the causal effects of SFHA designation. We specify and

estimate our equilibrium model of the housing market in Section 2.5. Section 2.6 simulates distributions of development and prices and discusses welfare under factual and counterfactual policies. We conclude by considering directions for future work and implications for policy.

2.2 Institutional Background

2.2.1 The National Flood Insurance Program and Special Flood Hazard Areas

Congress established the National Flood Insurance Program (NFIP) in 1968 in response to high flood losses and a perception that lackluster local regulation permitted excessive construction in high-risk areas (Burby, 2001). Today, the NFIP, administered by the Federal Emergency Management Authority (FEMA), remains the primary provider of flood risk protection and regulator of floodplain development in the United States. The NFIP underwrites over 90% of flood insurance policies, creates the most widely-used measures of flood risk through its flood mapping process, and sets construction standards for buildings in areas mapped as high risk (Kousky et al., 2018a).

After the NFIP was established in 1968, the program was rolled out to communities throughout the country in the late 1970s and 1980s.[4] When a community joined the NFIP, it obtained a Flood Insurance Rate Map (FIRM), produced by the NFIP using a hydrological study. After the FIRM was produced, buildings in specific areas had to comply with regulations, and flood insurance became available to homeowners.

In both the initial FIRM and subsequent, updated flood maps, the most important distinction for both insurance policies and floodplain regulation is between areas that are determined to be high-risk, known as Special Flood Hazard Areas (SFHAs), and those that are not. In this paper, we will refer to the SFHA as the "flood zone".[5] All new construction and substantial home improvements in the flood zone must comply with building regulations that require that a home's lowest floor lie above the Base Flood Elevation (BFE).[6] In the flood zone, some homeowners face a flood insurance mandate and all homeowners face higher flood insurance prices for otherwise-equivalent houses.[7] Though the insurance mandate is loosely enforced, approximately 50% of homeowners in the flood zone hold a flood insurance policy, compared to 2% outside of the flood zone (Bradt et al., 2021). In Florida, annual flood insurance premiums inside the flood zone cost twice as much as outside the flood zone: $855 compared to $435.

Throughout the United States, 10% of land[8] and 6% of properties are in the flood zone (First Street Foundation, 2020). Due to both climate change and population growth, the share of the US population at a level of risk that triggers SFHA classification is expected to rise from 13% to 15% by 2050 (Wing

[4]Communities are geographic units specific to the NFIP. They are generally municipalities or unincorporated areas of a county.

[5]In the parlance of the NFIP, a "flood zone" denotes a category of flood risk that determines insurance premium price (e.g. V, A, X, etc.). With apologies to FEMA, we reclaim this phrase with the hope of improving readability for a general audience.

[6]While popular images of elevated houses commonly show those on posts or piles, this adaptation tends to appear only in close proximity to the coast, where wave action can destroy walls. In the mostly inland areas we study, enclosed elevated foundations are more common. This approach allows garages and unfinished basements to be constructed at ground level. See Figure A-1 for an example of a house with an elevated foundation.

[7]Homeowners with federally-backed mortgages are legally required to purchase flood insurance.

[8]Authors' calculations using 2017 flood maps.

et al., 2018). This makes flood-zone-induced building requirements one of the most common forms of zoning regulation in the U.S., comparable to minimum lot area requirements, which apply to an estimated 16% of single-family homes (Song, 2021).

2.2.2 The Flood Mapping Process

Our spatial regression discontinuity approach relies on the assumption that flood zone delineation is a coarsening of a continuous measure of flood risk and does not follow the contours of true discontinuities in flood risk or other amenities. The validity of this assumption relies on the details of the mapping technology. In our specific context, there is substantial scope for imprecision in the historic boundaries we exploit.

The accuracy of a flood map depends on both the accuracy of the estimates of land elevation and the accuracy of the hydraulic model which simulates the amount of excess water in a flood event (National Research Council, 2009). Historically, engineers estimated land elevation based on US Geological Service contour lines, which suffer absolute elevation error on the order of meters.[9] After floodwater heights have been mapped, the floodplain is delineated by transforming vertical flood elevation profiles into horizontal floodplain boundaries. Because the same elevation of floodwaters yields a much wider floodplain in flat than steep areas, the floodplain boundary delineation is four to five times more uncertain in flat areas, such as Florida, compared to hillier areas (National Research Council, 2009). The floodplains of inland Florida are particularly uncertain since their drainage is dominated by shallow water flow, an atypical landscape for which FEMA does not specify hydrology and hydraulics guidelines.

After the construction of the initial flood map, FEMA is required to update these maps every 5 years to account for improved mapping technology and changes in development that may impact flood risk (National Research Council, 2009). In practice they are often updated much less frequently: as of 2017, more than 50% of maps were more than 5 years old (U. S. Office of Inspector General, 2017). In between official remapping cycles, property owners can request map amendments to correct inaccuracies (National Research Council, 2009) or petition for a map carve-out if homeowners have physically changed the land elevation (e.g. by adding dirt, called "fill"). According to our conversation with a floodplain manager, in the early years of the program the scale of paper maps meant that fill-based carve-outs of the flood zone had to be at least 6 acres. Because of this requirement, most houses did not find it cost-effective to pursue a carve-out. More recently, the adoption of digital maps has enabled these carve-outs at a smaller scale, and they have subsequently become more common.

2.3 Setting and Data

Our empirical context is the state of Florida, one of the most flood-prone and populous states in the United States. This makes it an ideal setting to study how floodplain regulation impacts housing markets and disaster damages. Nearly 50% of land and 19% of homes in Florida are located in the flood zone, underscoring the relevance of this form of regulation for real estate development across the state. (First Street Foundation,

[9] Today, LiDAR technology has improved the accuracy of land elevation models. More-powerful computing has also improved the precision of hydraulic modeling over time.

2020). Florida alone accounts for 35% of the nation's NFIP policies (Lingle and Kousky, 2018).[10] We bring together four primary sources of data to conduct our analysis.

Digitized historic and current flood maps Our analysis is organized around archival scans of early flood maps that we digitized for parts of eleven counties.[11] We aimed to collect the first Flood Insurance Rate Maps (FIRMs) ever drawn. In a few instances, constraints on the availability or formatting of these first maps made this impossible. In these cases, we were able to digitize maps that were drawn only a few years later. All but two of the 120 panels we digitized became effective between 1977 and 1984. Appendix A.2.1 describes the sample selection process in more detail.

Figure 2-1a presents an excerpt of these digitized flood maps. Figure 2-1b shows the geographic coverage of our digitized sample. While budget constraints prohibited digitizing the entire state of Florida, we are able to obtain good coverage of most major population centers. Table 2.1 illustrates that our sample covers 10.5% of the land mass in Florida, but 14% of all homes, reflecting the fact that our digitized areas are more developed and populous than average.

We pair our newly-digitized historic flood maps with snapshots of flood maps for the whole state from 1996 and 2017. In our digitized counties, 32% of land is in the flood zone.

Satellite-Derived Land Use Data Figure 2-1a demonstrates that the floodplain distinctions are detailed, necessitating spatially granular data on land use to study outcomes on either side of the boundary. We use two datasets to measure land use at two points in time. The first is US Geological Survey data on land use patterns contemporaneous to the time the original maps were drawn. This dataset consists of a 30x30 meter raster describing land use and land cover as belonging to one of nine mutually exclusive meta-categories, including urban/built-up land, agriculture, wetland, and water.[12] The categories were determined based on high-altitude photographs taken between 1971 and 1982 (1976 is the median and mode image date). We define "developed" land in this data as land falling into the "urban/built-up" category, which includes land used for residential, commercial, industrial, or transportation purposes. For current land use, we employ the National Land Cover Database (NLCD) from 2016, which classifies Landsat remote sensing imagery into similar categories of land cover, also in a 30x30 meter grid. Our main category of interest, "developed", indicates land that is covered by a mixture of constructed materials and mostly-lawn-grass vegetation.

Table 2.1 panel A presents land use summary statistics for the state of Florida and our digitized subsample. Commensurate with Florida's population boom between 1980 and 2020,[13] Table 2.1 illustrates that development increased substantially both statewide (2.6x) and in our sample of interest (2.5x).

[10] See https://nfipservices.floodsmart.gov//reports-flood-insurance-data for details.

[11] These archival scans were downloaded from FEMA's Map Service Center https://msc.fema.gov/portal/advanceSearch. In order to maximize power, we prioritized areas with substantial new development over the last 40 years. Our estimates on development when expressed in levels may therefore generalize less well to other settings, but this choice will not affect results expressed as a percentage of new development.

[12] Across Florida, the median number of raster grid cells per census tract is about 4900.

[13] Between 1980 and 2020, Florida's population more than doubled from 9.75 million to 21.5 million (US Census Bureau, 2022).

Parcel Characteristics Data from the Florida Department of Revenue property tax records from 2005 to 2020 provide detailed information about structures, including sales prices and parcel outline and geolocation. We precisely geolocate the exact location of any buildings on each parcel using Microsoft's open-source building footprints dataset.[14] Table 2.1 panel C summarizes average home prices statewide and in our sample of interest.

We supplement our data from the Florida Department of Revenue with data from NFIP claims and policies from 2010 to 2018 to provide information about elevation, policy cost, and flood insurance damages for insured structures. We also obtain historical data on home prices from the 1980 Census (Manson et al., 2021).

Flood Risk Model To assess the risk profile of development across policy regimes, we draw on spatially-granular estimates of flood risk from a third-party hydrological model. This model is produced by the First Street Foundation, a nonprofit organization devoted to quantifying and communicating climate risks. First Street aims to improve on government-issued risk assessments, which have been criticized for being out-of-date and inaccurate (Keller et al., 2017; Brannon and Blask, 2017; Frank, 2020b; Wing et al., 2022). First Street takes into account sources of flooding that NFIP maps ignore (e.g. rainfall), provides estimates for areas that FEMA had not been able to survey, and accounts for sea level rise due to climate change (First Street Foundation, 2020). Although First Street's model does not employ the "gold standard" of surveying that FEMA uses in the highest-risk locations, their validation exercises have achieved 80-90% flood extent similarity with historical observations and they are considered to "fus[e] the accuracy of local studies with the spatial continuity of large-scale models" (Wing et al., 2022). Nationally, First Street's model estimates that NFIP flood maps identify only 60% of areas that face a 1% chance of flooding every year (First Street Foundation, 2020). In Florida this discrepancy is smaller: only about 10% more houses face substantial risk under the First Street model than the current flood maps indicate. However, First Street and FEMA disagree about the exact location of this risk. Appendix Table A.1 tabulates the discrepancies between FEMA's flood maps and the First Street model in our sample, showing that more than one-fifth of parcels are categorized differently by FEMA and First Street.

2.4 Causal Evidence on the Effects of Floodplain Designation

We begin our analysis by describing the effects of flood zone designation on flood risk, which could occur both by shifting construction away from high risk areas and by mandating elevation in those risky areas. In this section, we employ a spatial regression discontinuity design around the regulatory boundary to study the policy's impact on the location of new construction, and perform an event study around the introduction of the policy to study its effect on house elevation and the consequent effects on flood damages.

[14] This dataset is also derived from satellite imagery, mostly captured in 2019. See https://github.com/microsoft/USBuildingFootprints for more details.

2.4.1 Boundary Analysis of the Effect of Floodplain Regulation on Construction and Prices

Our spatial regression discontinuity design compares current development and house prices on either side of the historic regulatory boundary to investigate the extent to which floodplain regulations reduce construction in risky areas. We also use our boundary discontinuity to compare differences in home prices in regulated versus unregulated areas. In the context of our model, these two equilibrium points – prices and quantities inside and outside the regulated areas – will later allow us to estimate the costs of regulation by revealed preference.

Empirical Strategy

Estimating the effect of flood zone designation on new construction presents two challenges. First, flood zone designation could be correlated with unobserved amenities, such as coastal access or views. Indeed, Table 2.1 indicates that land inside the flood zone is more likely to be water or wetlands and is closer to the coast. Second, flood zone designation may be endogenous to real estate construction, as the mapping process allows homeowners to deregulate parts of their properties by petitioning for map corrections or "filling" in dirt to elevate the land. Inside the flood zone, homeowners who are correctly mapped have an incentive to elevate their house to "escape" the flood zone. Homeowners who were incorrectly mapped have an incentive to petition FEMA to correct a mistake that overstates a home's risk. Meanwhile, owners of undeveloped land face no such incentives. Appendix Table A.2 shows direct evidence of such reverse causality: land that was developed as of 2004 is more likely to be remapped out of a floodplain in the next map revision than land that was undeveloped. This endogenous amendment process would lead to a mechanical negative correlation between development status and flood zone status that is unrelated to the causal effect of interest.

We address these two challenges with a spatial regression discontinuity design that leverages the first flood maps drawn in the late 1970s and early 1980s. The historic maps address concerns about reverse causality, since homeowner petitions and reactive adaptations were not reflected in the original maps: amendments and endogenous adaptation happen only after the maps are drawn. The regression discontinuity addresses omitted variables bias by leveraging the coarse classification of the flood zone and the assumption that unobservable characteristics of the land evolve smoothly through the historical flood zone boundary. We probe this identifying assumption by examining land use outcomes before or contemporaneous to the drawing of these initial flood maps.

We estimate our boundary discontinuity design by examining how outcomes at each 30x30m pixel vary as a function of the distance to the flood zone boundary. Formally, we estimate the following model:

$$y_i = \beta 1\{d_i > 0\} + f(d_i) + \gamma_{j(i)} + \epsilon_i \tag{2.1}$$

where y_i is a characteristic of pixel i and d_i is the perpendicular distance from each pixel i to the nearest flood zone boundary, with positive values indicating that the pixel is located inside the flood zone. Our coefficient of interest is β, the magnitude of the discontinuity at the boundary. We allow characteristics y_i to vary flexibly on either side of the boundary. In our baseline specification, we specify $f(d_i)$ as a fourth-order

polynomial that is allowed to differ on either side of the boundary. Results are similar under alternative local linear specifications. We cluster our standard errors at the census tract level to allow for spatial correlation in the error term. Finally, since our boundaries do not have natural segments, we include census tract fixed effects $\gamma_{j(i)}$ as a substitute for boundary fixed effects.

We estimate equation 2.1 on land within 2,000 feet (0.38 miles) of a flood zone boundary. Following previous work, we exclude boundaries that trace a body of water (Dell, 2010). Columns 4 and 5 of Table 2.1 present summary statistics for our boundary estimation sample: land close to a boundary is more developed and has slightly higher home prices than areas further from the boundary. Appendix Figure A-2 plots a histogram of the number of pixels in our estimation sample across distance-to-boundary bins.

Results

We discuss our results in the context of an *intent-to-treat* framework: the treatment of interest is the initial flood zone designation, which may evolve over time. Appendix Figure A-3 documents the evolution of the relationship between initial designation and floodplain status over time.

Exogeneity of boundaries To validate our identifying assumptions, we check for smoothness in land use through the regulatory boundaries prior to flood zone designation.[15] If preexisting amenities differed discontinuously across the boundary, or if boundaries were drawn around the contours of existing development, we would observe discontinuous patterns in development around the flood zone boundary. We test this by estimating Equation 2.1 with y_i equal to pre-period development. Figure 2-2a shows that pre-period development is smooth across the boundary, and the estimated coefficient, reported in Table 2.2, is economically small and statistically insignificant.[16] This test supports our hypothesis that the institutional details of the initial mapping process provide a compelling setting to conduct a boundary discontinuity analysis.

Development falls in the Flood Zone By 2016, we see a sharp discontinuity in development at the SFHA boundary, as shown in Figure 2-2b. Table 2.2 reports the estimated level shift at the boundary ($\hat{\beta}$), which indicates that land just inside the SFHA is 4.2 percentage points less likely to be developed than land just outside the SFHA.[17] This effect is substantial: it is 9% of the outside-SFHA mean level of development and it represents an 18% reduction as a share of total new development occurring between 1980 and 2016 just outside the SFHA.[18] The effect is driven by single family homes, which make up the majority of residences (87%) in our sample. The magnitude of the policy's negative effect indicates the potential for both substantial

[15] Our land use data was collected via aerial photographs between 1971 and 1982, while the flood maps were drawn between 1977 and 1984. Most aerial photographs were taken during or before 1976, before any of the maps were drawn. While it is possible that some aerial photographs were taken after the maps had been drawn, we will interpret these land use outcomes as a pre-period. We do not think this is particularly problematic, as land use evolves slowly and the worst-case scenario is that the photographs were taken five years after the drawing of the map.

[16] Appendix Figure A-4 and Table A.3 demonstrate smoothness in other pre-period land use outcomes. Note that in our alternative local-linear specification, the coefficient continues to be economically small but does become statistically significant.

[17] Figure 2-2b documents a striking pattern of land use on either side of the SFHA boundary. Moving away from the boundary further into the SFHA, development decreases sharply for about 400 feet before leveling out. Because this pattern is asymmetric, we do not believe it is driven by measurement error. Instead, it may be driven by a higher prevalence of wetlands deep inside the SFHA, which is also apparent in the pre-period land use (see Appendix Figure A-4).

[18] Table 2.2 Panel C shows that these results are robust to alternative definitions of development, including the share of land covered by a building footprint.

reductions in flood damages via reduced building in risky areas and costs to either developers or consumers that yield this behavioral response.

Prices increase in the Flood Zone The effects of floodplain regulation on house price are ex-ante ambiguous. On the one hand, by suppressing demand for regulated houses, flood zone designation could push prices down. On the other hand, by requiring suppliers to employ costlier construction methods, flood zone designation could push prices up. Figure 2-3, which plots the estimated coefficients for the sales price outcome, indicates that supply must shift inward substantially as a result of the policy: though we observe a large decrease in development, prices are non-decreasing through the boundary. In fact, our point estimates suggest that prices are 6% higher inside the flood zone, indicating that the construction costs imposed in the floodplain dominate any negative demand effect driven by mandatory flood insurance, higher flood insurance prices, or any salience or risk perception effects of living in a flood zone. These impacts are again driven by single family residential homes (Figure 2-3b).[19]

Our evidence suggests that this positive price effect should indeed be interpreted as an inward shift in supply. We rule out compositional changes in housing characteristics, including house elevation, as drivers of our result. In Appendix Table A.4, we report estimates using sales prices residualized of a rich set of observable characteristics of each house and show that the positive price effects persist. We also see similar effects on prices for homes built before and after the introduction of the flood zone regulations. We provide additional evidence to rule out changes in amenities driving a positive price effect: our results are robust to excluding areas close to the coast (column 5 of Table 2.2). Figure 2-3c shows that vacant land prices decrease by 10% across the flood zone boundary, a pattern that would not be consistent with increases in amenities.

Robustness Columns 3 and 4 of Table 2.2 illustrate the robustness of our results to alternative specifications to ensure that our results are not driven by specific choices in our estimation framework. Appendix Figure A-5 shows corresponding figures. We observe similar effect sizes on both quantities developed and prices if we estimate a linear function of the running variable estimated on a narrower bandwidth (500 feet) or if we estimate a local linear regression instead of a fourth order polynomial using the MSE-optimal bandwidth proposed by Calonico et al. (2014).

2.4.2 Event Study of the Effect of Floodplain Regulation on Elevation, Flood Damages, and Prices

The preceding results indicate that regulations impose costs on developers that shift supply inward. In this section, we investigate whether these regulations also reduce damages on the intensive margin of elevation.

[19]This finding contrasts with recent work that has found flood zone designation decreases house prices, e.g. Hino and Burke (2021). That work exploits map updates to achieve identification, and therefore primarily captures short-term demand effects. In our context, we study the effect of flood zone designation over the course of 40 years. This long-run setting allows supply to respond to mandatory building codes. An inward shift of the supply curve would reduce the quantity of houses built in the floodplain over that time, and could counteract the demand-driven price decrease.

Empirical Strategy

Following Wagner (2021), to estimate the effect of elevation we exploit the fact that a community had to adopt flood-safe building standards at the time it enrolled in the NFIP. Newly-built houses were required to be elevated to the level of the 100-year-flood just after a community enrolled in the NFIP — but not before. This suggests an event-study design in which we regress our outcomes of interest against the year a house was built relative to enrollment, within the flood zone:

$$y_i = \Sigma_r \beta_r 1\{r_i = r\} + \gamma_{j(i)} + \varepsilon_i \tag{2.2}$$

where y_i is an outcome of interest (elevation, insurance payouts, or sale price); r_i indicates the construction year relative to house i's year of enrollment; and $\gamma_{j(i)}$ is a set of census tract fixed effects.[20] We scale payouts by amount of purchased insurance coverage; since individual claims cannot be linked to individual policies we therefore estimate this model at the census-tract-by-flood-zone-by-relative-year-built level.[21] The sample for this analysis includes all counties in Florida. We cluster standard errors at the census tract level. In order to increase power, we also estimate a binned specification: we restrict to houses built within 10 years of the date a community joined the NFIP and estimate the following model, again clustering standard errors at the census tract level:

$$y_i = \alpha + \beta Post_i + \nu r_i + \eta r_i Post_i + \gamma_{j(i)} + \varepsilon_i \tag{2.3}$$

where $Post_i$ indicates being constructed in or after the year of NFIP enrollment ($r_i \geq 0$). Under the assumption that the year of construction was not manipulated, β indicates the causal effect of building code regulations on our outcomes of interest.[22]

Results

In Figure 2-4 we present the coefficient estimates on relative year from the event study specification (Equation 2.2). Variable means and regression coefficients from the binned specification are presented in Table 2.3. Table 2.3 also shows the results from a difference-in-difference specification that compares trends in houses in the floodplain to houses outside the floodplain. This specification, which is described in Appendix A.2.3 and graphed in Appendix Figure A-6, yields similar results.

[20] Census tracts are smaller than communities.

[21] We estimate these models on two datasets. The first, derived from NFIP claims and policy data, describes elevation status, insurance payouts, and policy cost between 2010 and 2018. The second dataset describes house sale prices by census tract, flood zone, and relative year built. Appendix A.2.3 describes the construction of these datasets in more detail. Appendix Table A.5 presents summary statistics for the datasets used in this analysis.

[22] Our results indicate smooth trends in elevation prior to the year of construction. We do not find jumps in the year just before NFIP enrollment, indicating that developer manipulation of construction year is not a concern.

Building regulations increase elevation and reduce flood damages Figure 2-4a shows that building regulations increase the share of homes built above the modern-day Base Flood Elevation (BFE) from 60% to 90%. We draw two conclusions: first, regulations increase elevation substantially, and second, a large share of homes were already built above the BFE. Our results suggest our measurement of pre-period elevation is most likely driven by areas where the BFE is sufficiently low that homes are built above it in standard construction practice. We arrive at this conclusion using two pieces of evidence. As we document below, consumers are unwilling to pay for elevation, meaning that costly elevation will not be provided absent regulation. Second, we find that the pre-period elevated homes are concentrated in areas with low flood risk, and hence low Base Flood Elevations (Appendix Figure A-8).

Figure 2-4b shows that the introduction of building regulations causes insurance payouts to fall by $1.60 per $1000 of coverage. At average coverage levels, this indicates that the increase in elevation mandated by the regulations does indeed generate social value in damage reduction across all houses in the flood zone of about 3% of the average home value (using a 5% discount rate).

House prices are unchanged despite substantial changes in damages Despite their higher social value, the prices of elevated houses are not significantly different from those of non-elevated houses. The point estimate of the increase in house price for elevated houses is -0.7%, or -$1400. Appendix Figure A-7 shows that this result is robust to residualizing the sales price against fourth-degree polynomials in parcel size and total living area, and fixed effects for county-by-month-of-sale and year of construction. Appendix A.2.3 shows in a stylized model that a zero effect on price implies zero willingness-to-pay for reduced risk if all other consumer-welfare-relevant aspects of houses remain constant. The null effect we find on price is not driven by a failure of insurance prices to reflect changes in risk, as premiums for elevated homes incorporate 75% of the reduction in risk (Figure 2-4c).

This result confirms existing work documenting a wedge between social and private valuations of flood risk in our setting (Gallagher, 2014; Kousky et al., 2018b; Bakkensen and Barrage, 2021; Wagner, 2021; Mulder, 2021). The fact that safe houses provide no private value to consumers indicates the presence of frictions, including behavioral frictions such as risk misperception or myopia, or moral hazard from consumer expectations of government aid in case of disaster (or both). Regardless of the source, this wedge between private and social value will yield construction of inefficiently-risky houses, absent regulation.

2.5 An Equilibrium Model of the Housing Market

Our reduced-form results inform us about the policy's effect on both flood damages and regulatory costs, and show the potential for the policy to improve welfare. Floodplain regulation suppresses construction and increases elevation in high-risk areas. The policy therefore potentially decreases damages on both extensive and intensive margins of construction. However, our results also indicate that the costs of mandatory elevation are substantial, since flood zone designation substantially decreases new construction without a commensurate fall in price. Finally, although elevation reduces damages, consumers are not willing to pay more for elevated houses, suggesting a role for policy to restore efficient risk allocations.

While informative, these results alone are insufficient to quantify the total effect of the policy on either damages or overall welfare. The policy's effects depend critically on the location of counterfactual construction:

if the regulation shifts construction to equally-risky areas, damages will not fall. Counterfactual risk will also depend on the joint distribution of risk and amenities, since more-desirable locations will attract more counterfactual construction. Additionally, the large size of the regulated area — nearly one-third of all land in our digitized sample is currently designated as a flood zone — may introduce spillover effects on the unregulated locations. Higher demand and consequently higher prices outside the flood zone could mitigate the incentive to relocate and impose pecuniary externalities on unregulated consumers. Moreover, while our empirical results indicate that floodplain regulation shifts supply, we require a model to quantify the regulation's costs to developers and consumers. To account for counterfactual location and spillover effects, quantify the regulation's costs, and compare welfare across current and counterfactual policies, we specify and estimate a model of residential choice and real estate development.

Our model's key parameters of interest describe how the flood zone designation shifts the demand and supply curves inwards. These parameters capture the "direct effect" of the flood zone on demand and supply: they measure how consumers and developers trade off home prices and living or building in a regulated area. Both of our empirical exercises inform our model. We use our event study estimates that consumers are unwilling to pay for elevation to simplify the model: consumers choose only among differentiated locations and developers do not endogenously supply elevated homes. We use our cross-boundary changes in prices and quantities from our boundary discontinuity analysis as moments to estimate the model's key parameters. Using these estimated parameters, we then simulate the housing market to quantify the effect of factual and counterfactual policies on new construction, home prices, expected damages, and welfare.

2.5.1 Residential choice

Motivated by our result in Section 2.4.2 that consumers are indifferent between elevated and non-elevated homes, we model consumers as choosing among uniform homes across differentiated locations. We further assume that consumers do not privately value differences in flood risk when deciding between houses, consistent with prior work and with our results in Section 2.4.2. Each individual i makes a discrete choice of where to live within market m, which we take to be a county.[23] Locations are differentiated goods characterized by tract j and flood zone designation status z. Census tracts are small geographic units of analysis: the average county in Florida has 63 census tracts, each containing roughly 1,600 residential structures. Following the standard discrete choice framework of Berry et al. (1995), the indirect utility of individual i living in location jz is given by:

$$u_{ijz} = \alpha^D p_{jz} + \phi_C SFHA_z + \xi_{jz} + \varepsilon_{ijz} \tag{2.4}$$

where p_{jz} is the log price of housing in location jz, $\phi_C SFHA_z$ indicates financial and non- financial costs of living in an area designated as a flood zone, ξ_{jz} are unobserved amenities, and ε_{ijz} is an i.i.d. preference shock. Non-financial costs of living in a floodplain incorporated in $\phi_C SFHA_z$ can include the hassle cost of regulations, such as the requirement to document flood insurance purchase when initiating a mortgage, or anxiety around future flood risks that is introduced by the high-risk label.

Because consumers purchasing a house may face information frictions, for example underestimating a house's flood risk, we allow for the possibility that flood zone designation may debias consumers. Following Allcott

[23] In Florida, counties are large but tend to only contain one major city and commuting zone.

and Taubinsky (2015), we account for this by allowing flood zone designation to impact decision utility but not experienced utility. We model consumers as purchasing houses following a decision utility of

$$u_{ijz} = \alpha^D p_{jz} + \phi_C SFHA_z + \phi_B SFHA_z + \xi_{jz} + \varepsilon_{ijz} \tag{2.5}$$

where p_{jz} is the log price of housing and we have introduced $\phi_B SFHA_z$ as a potential channel for flood zone designation to affect beliefs about flood risk. Modeling the debiasing effect of the flood zone as an on-or-off label captures the fact that flood maps do not give any indication of the expected flood damages of a particular location; they simply indicate a coarse level of risk.

We cannot separately identify real costs ϕ_C and debiasing labels ϕ_B since both vary with floodplain status. We therefore combine the real cost term $\phi_C SFHA_Z$ and debiasing effect term $\phi_B SFHA_Z$ into an overall "floodplain effect" term $\phi SFHA_z$.

The model of consumer decision utility we take to the data is therefore

$$u_{ijz} = \alpha^D p_{jz} + \phi SFHA_z + \xi_{jz} + \varepsilon_{ijz} \tag{2.6}$$

where p_{jz} is the log price of housing in location jz,[24] $SFHA_z$ indicates whether z is regulated as a floodplain, ξ_{jz} are unobserved amenities, and ε_{ijz} is an i.i.d. preference shock. We assume ε_{ijz} is distributed according to an Type 1 Extreme Value Distribution.

Individuals choose the location jz that maximizes their idiosyncratic decision utility within the locations in market m. Because ε_{ijz} is distributed EVT1, the fraction of individuals in a county choosing to live in location jz is:

$$s_{jz} = \frac{\exp(\delta_{jz})}{\Sigma_{j' \in J_m, z' \in \{0,1\}} \exp(\delta_{j'z'})} \tag{2.7}$$

where $\delta_{jz} = \alpha^D p_{jz} + \phi SFHA_z + \xi_{jz}$ indicates the mean (non-idiosyncratic) utility for each tract-zone pair.

2.5.2 Housing supply

To flexibly capture heterogeneity in supply elasticities across locations and within locations across time, we model development in each location jz as the result of decisions of granular, heterogeneous landowners. Each tract-zone pair is composed of L_{jz} plots, each of which could either be developed into a house or used for some outside option (e.g. agriculture). The value of the outside option for plot g is denoted c_g (for opportunity cost) and distributed Normally with a mean and standard deviation that varies by Census tract j: $c_g \sim N(\mu_j, \sigma_j^2)$. Developers make static decisions about whether to develop at two points in time: before the regulations are imposed ($t = 0$) and after they are imposed ($t = 1$).

[24] Note that the price p_{jz} is for the bundle of housing that a consumer purchases, which includes both the structure and the land on which the structure is built.

The value of developing a house in period t depends on the (log) price p_{jz}^t for which it could sell, which varies by location and time, and the cost to build the house η_{jz}^t, which also varies by location and time and increases by a constant amount ψ when the house is elevated. Whether a house is elevated is determined exogenously and varies by tract and time period.

The development decision for an undeveloped plot g in time period t is given by

$$D_g^t = 1\{p_{jz}^t \geq c_g + \psi E_{jz}^t + \eta_{jz}^t\} \tag{2.8}$$

where $D_g^t = 1$ indicates that a plot of land is developed and E_{jz}^t indicates whether or not houses built in location jz are elevated in period t. The total share developed at the end of time $t = 1$ is therefore $\Phi\left(\frac{p_{jz}^1 - \psi E_{jz}^1 - \mu_j - \eta_{jz}^1}{\sigma_j}\right)$.

2.5.3 Equilibrium

In equilibrium, house prices and location decisions adjust so that the quantity of housing supplied in each location equals the number of individuals choosing to live there. Specifically: $L_{jz}\Phi\left(\frac{p_{jz}^1 - \psi E_{jz}^1 - \mu_j - \eta_{jz}^1}{\sigma_j}\right) = Q_{jz} = N_m s_{jz}$ where L_{jz} is the total amount of land in location jz, Q_{jz} is the quantity of developed land in jz, and N_m is the number of households in the market.

2.5.4 Estimation and Results

A key challenge for estimation arises from the fact that amenities ξ_{jz} and construction costs η_{jz} may be correlated with floodplain status $SFHA_z$.[25] The correlation between $SFHA_z$ and ξ_{jz} and η_{jz} occurs because flood zone status may be correlated with other amenities, like coastal access, or challenges to construction, like wetlands. We use a variation of our spatial discontinuity assumption to account for the endogeneity between flood zone status and unobserved amenities ξ_{jz} and construction costs η_{jz} to estimate regulatory costs ψ and ϕ. We describe the estimation procedure in more detail below.

Demand-side identification

We use the standard Berry inversion to obtain δ_{jz} from observed market shares:

$$ln(s_{jz}) - ln(s_{0m}) = \delta_{jz} = \alpha^D p_{jz} + \phi SFHA_z + \xi_{jz} \tag{2.9}$$

[25] Another important source of endogeneity is the correlation between p_{jz} and ξ_{jz} and η_{jz}, as locations with higher amenities and higher construction costs have higher prices in equilibrium. However, because we calibrate supply and demand elasticities rather than estimate them in-sample, we avoid this additional estimation challenge.

where s_{0m} indicates the market share of the arbitrary geography within each market that we have normalized to be utility 0. We construct the empirical market shares using

$$\hat{s}_{jz} = \frac{Q_{jz}^{2016}}{\sum_{j' \in J_m, z' \in \{0,1\}} Q_{j'z'}^{2016}} \tag{2.10}$$

where Q_{jz}^{2016} is the total amount of developed land in geography jz.

Simply estimating Equation 2.9 via OLS would yield a biased estimate of the regulatory demand cost ϕ because of the correlation between $SFHA_z$ and amenities ξ_{jz}. We address this challenge in our boundary discontinuity analysis by identifying the effect of flood zone designation in the limit on either side of the boundary. We therefore estimate our model of housing choice by matching the cross-boundary differences we document in the descriptive analysis. First, we restrict to a narrow window around the boundary (100 feet). Because even this narrow window does not capture differences at the limit, we then separate the unobserved amenity term ξ_{jz} into a nuisance term, which can be correlated with flood zone status, and a component of ξ_{jz} that is uncorrelated with flood zone status. We require that the price difference across boundaries net of the nuisance term match the price difference estimated in our spatial RD, denoted $\beta^{p,2016}$. Appendix A.2.5 specifies our moment conditions in more detail.

We calibrate the substitution elasticity α^D to values in the literature. In similar settings, authors estimating within-city location substitution elasticities have found values corresponding to our α^D in the range of -0.4 to -3.3 (Calder-Wang, 2021; Song, 2021).[26] The model in Song (2021) is of the most similar granularity to ours and estimates a value for α^D ranging from -0.9 to -0.99, so we calibrate α^D to be -1.[27]

Discussion This estimation strategy identifies the floodplain effect on demand (ϕ) close to the floodplain boundary, but assumes the cost of regulation is identical across all regulated areas. The floodplain boundary tends to have lower risk than areas deeper inside the floodplain. Thus, to the extent that the true floodplain effect on demand is larger in higher-risk areas (e.g. coastal areas subject to storm surge) ours will be an underestimate.

We have imposed the assumption that consumers do not respond to expected flood damages. If this assumption is incorrect, any increases in flood expenses to which consumers do react will load onto the flood zone term if they change discretely at the boundary (e.g. through insurance premiums) or onto the unobserved amenity terms ξ_{jz} if they do not change discretely at the boundary. Underestimating consumer responsiveness to risk would lead us to overstate the gains from regulation in our normative analysis, but would not impact any conclusions about the positive effects of the policy on construction, damages, or house prices.

Supply-side identification

We assume that the flood zone designation only affects supply by requiring houses to be elevated, so we employ a similar strategy to identify the elevation cost term ψ as with the demand cost term ϕ. We again

[26] To impute elasticities from values presented in Calder-Wang (2021), we use an average NYC apartment rental price of $1615 and neighborhood-specific average rental prices from 2019 from furmancenter.org.

[27] In Appendix Table A.7 we present a set of results in which we estimate α^D using variation in price in our sample, which yields a value of approximately -0.8.

restrict to land within 100 feet of the boundary and match the spatial RD estimates of the flood zone effect to the cross-boundary differences in share developed, net of the component of η_{jz} that is correlated with amenities. We match the cross-boundary difference in share developed in both the pre-period ($\beta^{q,1980}$) and the post-period ($\beta^{q,2016}$). Details on the moment conditions we use can be found in Appendix A.2.5.

We also construct moments to match recent estimates of tract-specific supply elasticities from Baum-Snow and Han (2022) to obtain estimates of supply parameters μ_j and σ_j. These supply elasticities are estimated in the same locations as our setting, and we view them as accurately identifying both the magnitude of and spatial heterogeneity in the slopes of the housing supply curves in our sample. Using these estimates allows us to capture realistic patterns of within-market heterogeneity in supply elasticities, e.g. that housing supply in urban, coastal areas is likely much more inelastic than inland, suburban or rural areas.

Results

After calibrating α^D and matching μ_j and σ_j to estimates from Baum-Snow and Han (2022), we estimate the flood zone cost terms ϕ and ψ jointly via the Generalized Method of Moments using the boundary discontinuity moments described above and detailed in Appendix A.2.5. Appendix A.2.6 includes more details about the data restrictions used for estimation.

Model fit While our structural error terms allow our model to achieve a perfect fit to the observed data, we can assess the model by investigating the extent to which we rely on structural error terms to achieve that fit. In Figure 2-5 we plot observed price in 1980 and 2016 against the modeled prices that would rationalize observed market shares in each year, if we omitted the structural error terms. This exercise does not use our estimated demand model at all; it assesses the role of supply structural error terms based only on our estimated supply parameters and our method of simulating an equilibrium. We see a strong correlation between model-generated and observed prices, indicating our supply curve is reasonable.

Parameter estimates Table 2.4 presents the parameter estimates for our baseline specification.[28] Standard errors from bootstrapping with 100 replications are given in parentheses. Large standard errors are due to small sample size, a result of the cuts made when restricting to census tracts with sufficient data. We focus on the point estimates. As expected, flood zone status is disliked by consumers (negative ϕ) and imposes costs on developers (positive ψ). On the demand side, the magnitude of the flood zone cost implies that consumers are willing to pay 23% more to avoid living in a floodplain and it costs developers 25% more in construction costs to build a compliant home. We arrive at these numbers by rationalizing the changes in quantity and price around the regulatory boundary with estimates of how consumers trade off home prices with other attributes (our calibrated α^D) and the elasticity of housing supply (which is determined by μ_j and σ_j).

The result that consumers are willing to pay 23% more to avoid floodplain regulations is at the high end of a range of recent estimates, which find floodplain discounts ranging from 1 to 28% (Indaco et al., 2019; Gibson and Mullins, 2020; Hino and Burke, 2021; Lee, 2022). This difference could be attributed to our time period and setting: residents living in flood-prone Florida today may be particularly sensitive to signals of

[28] Appendix Table A.7 presents parameter estimates under alternative calibrations of demand elasticity α^D.

risk as awareness of climate change grows (Bernstein et al., 2019). Nevertheless, a 23% premium on avoiding the floodplain far exceeds the risk-reducing benefits of relocation: the risk difference across boundaries is equivalent to just over 11% of house price, just half of the estimated demand effect. This large discount may arise from consumers' strong dislike of the bureaucratic burdens of complying with floodplain regulations, which are substantial.[29] Alternatively, the flood zone designation may cause consumers to over-update their beliefs about risk, as flood maps do not communicate risks in terms of expected damages.

Turning to supply, we estimate that regulations produce a shift in the supply curve that corresponds to a 25% increase in construction costs in the flood zone.[30] The magnitude of this effect is large, but within the plausible (albeit wide) range of estimates of the effect of building codes and zoning regulations on construction costs: from 5% to 42% (Listokin and Hattis, 2005; Emrath, 2021; Song, 2021). The wide variation reflects both differences in strategies to estimate regulatory costs and variation in the types of regulations imposed. Yet, our informal interviews indicated that a 25% increase in costs is reasonable. For example, the elevation requirement may necessitate a stem wall, which can add $100,000 to the cost of a new build. The results in Section 2.4.2, combined with estimates of average risk levels from our hydrological model, indicate that elevation reduces expected flood damages by 55% among compliers. It follows that elevating an average home in a flood zone reduces damages by approximately 11% of house value, so on average, the benefits of mandatory elevation are less than half the costs. However, this comparison masks substantial heterogeneity across locations within the flood zone: elevating homes in the highest-risk areas will generate social value, while elevating relatively-safe homes will incur net social costs.

Taking our results together, we conclude that the flood zone regulations substantially shift both demand and supply curves inwards. The large costs of regulation to both consumers and developers implied by our boundary discontinuity results indicate that the policy imposes substantial burdens on consumers living in and developers building in flood zones. Back-of-the-envelope comparisons of estimated costs to risk-reducing benefits reveal that the costs may equal or outweigh the benefits of the policy. We explore this possibility rigorously by quantifying the total benefits and costs of the regulation in the following section.

2.6 Model-Informed Estimates of The Effects of Floodplain Regulation

We use our model to investigate the impact of floodplain regulation by comparing observed outcomes to counterfactual outcomes simulating the absence of the policy. We also compare the effect of the observed policy to the effect of a counterfactual first-best policy that imposes corrective taxes in the spirit of Pigou (1920) on consumers that are equal to the present discounted value of expected damages in each tract-by-flood-zone.[31]

We compute counterfactual outcomes by searching for a price vector $\tilde{P}_{jz}^{2016}$ that equates housing supply and

[29] In particular, non-elevated houses inside the floodplain must be elevated if they undergo substantial renovation. In an interview, a Florida developer said this rule burdens floodplain homeowners, whose remodeling plans must either include unwanted and expensive elevation or be spread out over several years to avoid triggering the rule.

[30] Construction costs include both the land on which a building is built and labor and materials costs.

[31] This policy could potentially be implemented as mandatory flood insurance with actuarially-fair rates at the property level. Recent changes to the National Flood Insurance Program moved the program in this direction — from a very coarse to a more granular pricing scheme — though the prices likely will not grow to reflect the level of damages estimated by First Street Foundation (Flavelle, 2021).

demand:

$$L_{jz}\Phi\left(\frac{\ln\left(\tilde{P}_{jz}^{2016}\right)-\hat{\psi}E_{jz}^{2016}-\mu_j-\eta_{jz}^{2016}}{\sigma_j}\right)= \tag{2.11}$$
$$N_m\frac{\exp(\alpha^D\ln\left(\tilde{P}_{jz}^{2016}+\tau_{jz}\right)+\hat{\phi}SFHA_z+\xi_{jz})}{\Sigma_{j'\in J_m,z'\in\{0,1\}}\exp(\alpha^D\ln\left(\tilde{P}_{jz}^{2016}+\tau_{jz}\right)+\hat{\phi}SFHA_{z'}+\xi_{j'z'})}$$

given our calibrated and estimated parameters $(\alpha^D,\hat{\mu}_j,\hat{\sigma}_j,\hat{\phi},\hat{\psi})$, our recovered values of ξ_{jz} and η_{jz}, and counterfactual values for the flood zone designation $SFHA_z$ and taxes τ_{jz}.[32] Under the taxation policy, taxes vary with both location and elevation status. In this scenario, to capture the fact that consumers would demand and developers would endogenously supply elevated houses in some locations, we impose that locations elevate when it is socially-efficient to do so. That is, we assume that houses are elevated when the gains from lower tax rates on an elevated house exceed the elevation costs. This approach generates counterfactual prices and quantities of development in each of the locations jz. Our counterfactuals assume a closed city: we assume the 2016 population in each of our ten counties remains constant across counterfactual scenarios.[33]

First, we quantify the magnitude of flood-risk reduction each policy achieves – the intended goal. We also decompose the risk reduction into components due to relocation to safer areas versus elevation in risky areas. In addition to our model of housing supply and demand, this risk quantification exercise crucially relies on our hydrological-model-based estimates of flood risk by location and our estimates from Section 2.4.2 on the risk-reducing benefits of elevation.[34] Appendix A.2.7 describes how we compute expected damages in further detail.

Quantifying the risk-reducing benefits of each policy does not require a normative welfare framework. In Section 2.6.2, we will use our model normatively to quantify the costs of the regulation to consumers and to calculate the welfare effects of observed and counterfactual regulations.

2.6.1 Policy Effects on Flood Damages

Our quantification of each policy's effect on flood damages focuses on two main outcomes. First, we study the approximate number of houses in the regulated area (observed flood zones).[35] This analysis extrapolates the effect of our regression discontinuity away from the boundary to investigate the extent to which floodplain

[32] Under the observed policy, we set τ_{jz} equal to 0 to reflect the fact that the main driver of insurance premiums is flood zone designation, so consumers face no other prices that vary with flood risk.

[33] We estimate the parameters on 11 counties but encountered difficulties resolving the counterfactual equilibria for Miami-Dade county. Therefore, we omit it in these counterfactuals.

[34] We use this external measure of flood risk rather than the FEMA flood maps for a few reasons. First, the FEMA flood maps have received extensive criticism for being out-of-date and backwards-looking (Keller et al., 2017; Brannon and Blask, 2017; Frank, 2020b; Wing et al., 2022) and for failing to include certain important components of flood risk, e.g. pluvial (rainfall) risk. The First Street model incorporates climate change predictions as well as all major flood drivers in a novel peer-reviewed approach. Second, the First Street flood risk estimates provide granular estimates of the average annual loss for each parcel. First Street's estimates are increasingly used in economics research as an independent assessment of flood risk (Bradt et al., 2021; Mulder, 2021; Sastry, 2021).

[35] We use the approximation that one house equals 900 square meters of developed land. Each grid cell in our simulation is 900 square meters. In recent years, the size of a new lot in Florida has been about one-quarter acre, or 1,000 square meters (Morrell, 2018).

regulation reduces building in regulated areas market-wide. Because regulated areas are on average riskier than unregulated areas, shifting development from regulated to unregulated areas will likely reduce damages, but the magnitude is uncertain and depends on the joint distribution of amenities and flood risk. Our second primary outcome therefore quantifies the reduction in damages directly, including both relocation and elevation. To complete our analysis of the positive effects of each policy on housing markets, we also report total number of elevated houses and average home prices inside and outside of regulated areas.

We estimate the effect of floodplain regulation by setting $SFHA_z = 0$ and $\tau_{jz} = 0$ to simulate outcomes in the absence of any floodplain regulation. On the supply side, this amounts to eliminating all flood-zone-imposed building regulations: there is no requirement to elevate to a minimum level. On the demand side, we think of eliminating the existing flood zone regulations as a policy regime in which consumers receive no signals indicating higher risk and have no mandate to purchase insurance or other hassles of living in the flood zone. Under this policy regime, there are no differences in insurance prices across areas. We also estimate the effect of a counterfactual policy that both eliminates all flood zone regulations and also imposes corrective taxes equal to the expected damages by location and elevation status. In our model, this counterfactual yields socially-optimal development decisions.

Table 2.5 presents results. Relative to our no-flood-zone counterfactual, the observed flood zone designation policy reduces the number of homes built in regulated areas by 15.5%, reallocating over 117,000 houses. Note that this effect is smaller than our corresponding boundary discontinuity estimate (an 18% reduction in development). This is due to two facts. First, our model accounts for the market-wide distribution of unobserved amenities. At the boundary, amenities are identical, so substitution from regulated to unregulated areas is more likely. At the market level, amenities are correlated with regulation status, so substitution will be less elastic, making the boundary discontinuity estimate an over-estimate of the total effect. Second, because such a large share of Florida is at risk of flooding, regulations have market-wide price effects that undermine the effect of the policy. As Table 2.5 illustrates, prices outside of the regulated area increase as consumers substitute to unregulated areas.

The policy achieves its goal of reducing flood damages via both relocation and elevation. Overall, the policy reduces the present discounted value of damages by about $7,800 per newly-developed house, which amounts to a 60% reduction in damages. Relocation contributes about one-quarter of that effect, while the requirement to elevate houses contributes the remainder.[36] The policy's benefits in terms of reduced damages are substantial, both in absolute numbers (a decrease in $3.5 billion of expected damages per county), and in comparison to alternative policy instruments: the number of homes relocated exceeds the total number of houses removed from risky areas by the NFIP's home buyouts program across the entire nation over its lifetime (Frank, 2020a).

The status quo policy reduces damages by approximately the same amount as the corrective tax policy. However, the corrective tax policy achieves these same reductions in damages with fewer distortions to housing decisions. Figure 2-6 illustrates the mechanism driving these differences. Relative to the status quo policy, the tax policy allows more construction and damages in moderately-risky areas while reducing construction and damages in the risky tail. It achieves this result by imposing regulatory costs that are proportional to flood risk, rather than imposing regulations uniformly across the designated flood zone.

[36] We calculate our decomposition based on percent rather than level reductions in damages, since changing either elevation or location behavior first dramatically changes the base damages from which we compute the effect of the other behavior in levels.

These findings suggest that although the current floodplain regulation policy generates its intended social benefits in the form of substantially-reduced damages — and even achieves the socially-optimal level of those reductions — those same benefits could be achieved with smaller distortions under a more-targeted policy. These distortions may be significant. Thus, a key question is to what extent the costs of these housing market distortions offset the policy's benefits.

2.6.2 Policy Effects on Welfare

In this section, we develop a welfare framework that allows us to quantify the policy's costs and compare these costs against the estimated benefits discussed in Section 2.6.1. Because it is possible that the flood zone both imposes real costs and debiases consumers, we must distinguish between decision and experienced utility when evaluating consumer welfare. As discussed in Section 2.5, the experienced utility of a consumer choosing housing in location jz is given by:

$$u_{ijz} = \alpha^D \ln(P_{jz}) + \phi_C SFHA_z + \xi_{jz} + \varepsilon_{ijz}$$

Though government policies such as subsidized insurance premiums may shield consumers from some flood risk, the social planner values all flood costs. We therefore define the social welfare function for consumers — that is, consumer surplus augmented by the social value of externalities and internalities and any government revenue — as:

$$\Sigma_{jz} N_{jz} \left(\alpha^D \ln(P_{jz} + D_{jz} + R) + \phi_C SFHA_z + \xi_{jz} + \varepsilon_{ijz} \right) \tag{2.12}$$

where P_{jz} is a location's price in levels, D_{jz} is the PDV of a location's expected damages, and R is any revenue generated by the policy that is lump-sum transferred back to all consumers. Recall that our estimated ϕ equals $\phi_C + \phi_B$, that is, it captures the combination of real costs imposed by the regulation (ϕ_C) and a debiasing effect that does not enter social welfare (ϕ_B). Because we only estimate the combined effect ϕ, we consider the spectrum of possible welfare effects by calculating welfare under two polar cases: (a) ϕ is 100% experienced costs or (b) ϕ contains only insurance premiums plus other terms that do not affect welfare.

Following Train (2009), we compute per-person consumer surplus in each market as $CS_i = \frac{-\bar{P}}{\alpha^D} \ln \Sigma_{j,z} \exp(\delta_{jz})$, where $\bar{P}$ is the average house price in levels. In the fully-welfare-relevant case we take $\delta_{jz} = \alpha^D p_{jz} + \phi SFHA_z + \xi_{jz}$; in the debiasing-plus-premiums case we take $\delta_{jz} = \alpha^D p_{jz} + \phi_F SFHA_z + \xi_{jz}$, where ϕ_F reflects our estimate that the PDV of insurance premiums amounts to about 5% of average house price.

We account for the impact of each policy on producer surplus by computing the change in producer surplus relative to the unregulated baseline. Specifically, for each observation we compute $Q\Delta P + \frac{\Delta Q \Delta P}{2}$, where Q is the number of newly-constructed houses and ΔP is the price change relative to no-flood-zone in levels, plus, in the case where elevation changes as a result of the policy, the cost of elevation in dollars (approximated as ψP_{jz}).[37] We then define total social welfare as social consumer welfare plus producer surplus.

[37] We take this approach rather than calculating producer surplus directly from the model in order to avoid complications introduced when converting between log dollars and level dollars. This approximation does not impact our overall welfare or policy conclusions.

We report results of these welfare calculations in Table 2.6 by comparing total social welfare under the observed regulation and the counterfactual corrective tax policy to the unregulated benchmark. Even if the floodplain demand cost ϕ mostly captures a debiasing channel rather than imposing costs on consumers, we estimate that the policy does not substantially improve social welfare: at most it increases welfare by $140 million per county. This is just over 10% of the benefits that could be generated under the corrective tax policy. The poor performance of the status quo policy is further reinforced if the policy introduces non-financial costs to consumers rather than debiasing them. We ignore any potential effects of each policy on net government revenue through the channel of subsidized premiums or disaster aid; accounting for the cost of public funds would further strengthen the performance of the corrective tax policy relative to the status quo.

Our results indicate that current policy achieves at best a small fraction of the maximum possible benefits of regulation. Because of the wedge between the private and social value of flood safety, risk would be inefficiently-high in the absence of any policy intervention. Indeed, Table 2.6 shows that a well-targeted policy intervention can substantially improve welfare, by more than $2,700 per newly-developed house. However, because existing policy applies strong incentives equally to more- and less-risky areas, the policy may yield too little construction in moderately-risky areas, while still permitting too much construction in the riskiest locations. A corrective tax policy could improve upon the existing policy by targeting strong elevation and relocation incentives to the riskiest areas and foregoing onerous regulation in safer areas.

Our results highlight an important policy lesson. Some proposals (e.g. home buyouts (Frank, 2020a)) focus purely on the savings that could be achieved by reducing risk. However, equivalent risk reductions can entail substantially different costs. A strong but poorly-targeted, undifferentiated policy may overcorrect risk in some areas, imposing costs on developers and consumers that are not warranted by the benefits. Our results underscore the gains to transitioning from a blunt attribute-based regulation to a more flexible policy instrument. While the NFIP is indeed transitioning to a more flexible pricing system (Risk Rating 2.0), consumers may be insensitive to premium increases without further efforts to increase salience. NFIP premiums will also need to grow substantially to accurately reflect risk: expected damages are currently estimated to exceed current premiums in Florida by a factor of four (First Street Foundation, 2021). Moreover, the costly elevation mandate is still in place for all areas where FEMA calculates there is a more than 1% chance of flooding each year. In many less-risky parts of these areas, our estimates indicate elevation is not cost-effective and should not be required. With appropriate changes, however, we find that floodplain regulation may be able to achieve substantial welfare gains.

2.7 Conclusion

For over 40 years, federal floodplain regulations have influenced housing markets, with limited evidence on either their costs or benefits. This paper combines a spatial regression discontinuity analysis of floodplain boundaries, an event study of the introduction of flood-safe building codes, and a spatially-granular model of the housing market to investigate the policy's equilibrium effects on the location of construction, housing prices, expected damages, and social welfare.

We find that local to the regulatory boundary, floodplain regulation decreases construction and increases house prices. We also find that flood-safe building codes reduce flood damages among all houses by about 3%

of house value, but these reductions are not privately valued by consumers.[38] Using our model to interpret our reduced-form results, we find that floodplain regulation in Florida reduces damages by 60%, with one-quarter of the effect accruing through relocation and the remaining three-quarters through elevation. Though the policy achieves substantial benefits in terms of flood damage reductions, it is poorly-targeted: in some areas, flood zone designation induces responses whose costs exceed their benefits, and in others, it does not go far enough. Because of this, at best the policy only achieves about one-tenth of the possible regulatory gains, and at worst may actually harm welfare.

Our findings have important implications for policies to promote the resilience of cities in the face of sea level rise and other climate-change-induced increases in flooding. They suggest that current policy instruments may perform substantially worse than flexible instruments that allow consumers to live in risky areas if they are willing to pay the full social cost of doing so.

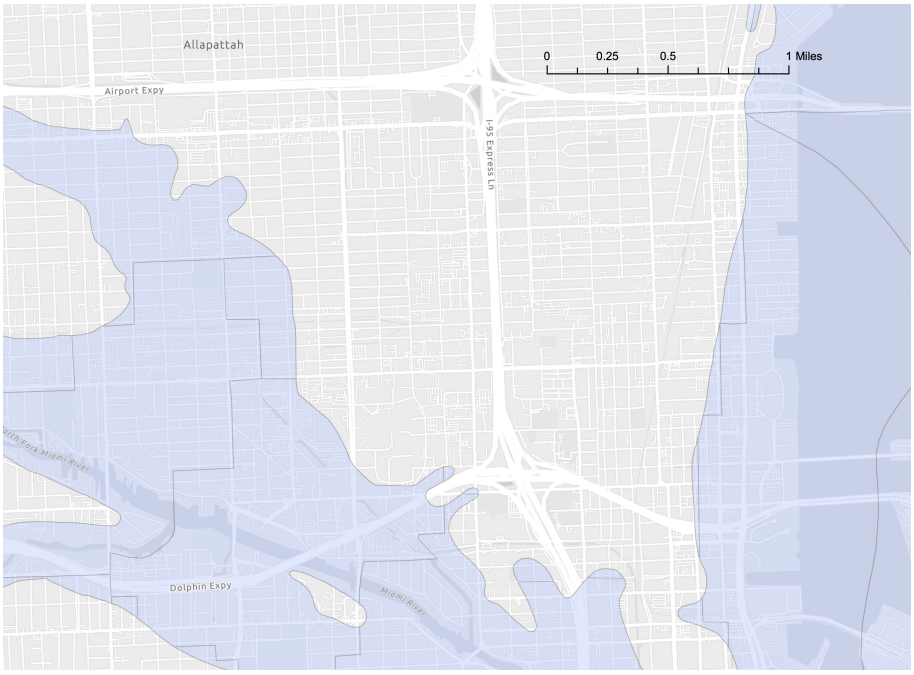
We leave a few areas open for future work. Though we model flood risk as static after 2050, in reality flood risk could increase further over time. Regulating durable construction with evolving flood risk may raise interesting dynamic considerations. Furthermore, although we provide suggestive evidence of frictions in the market for flood safety, we leave the challenge of identifying the welfare costs of each separate friction to future work.

[38] Elevation reduces damages for a given elevated house inside the flood zone by about 11% of house value.

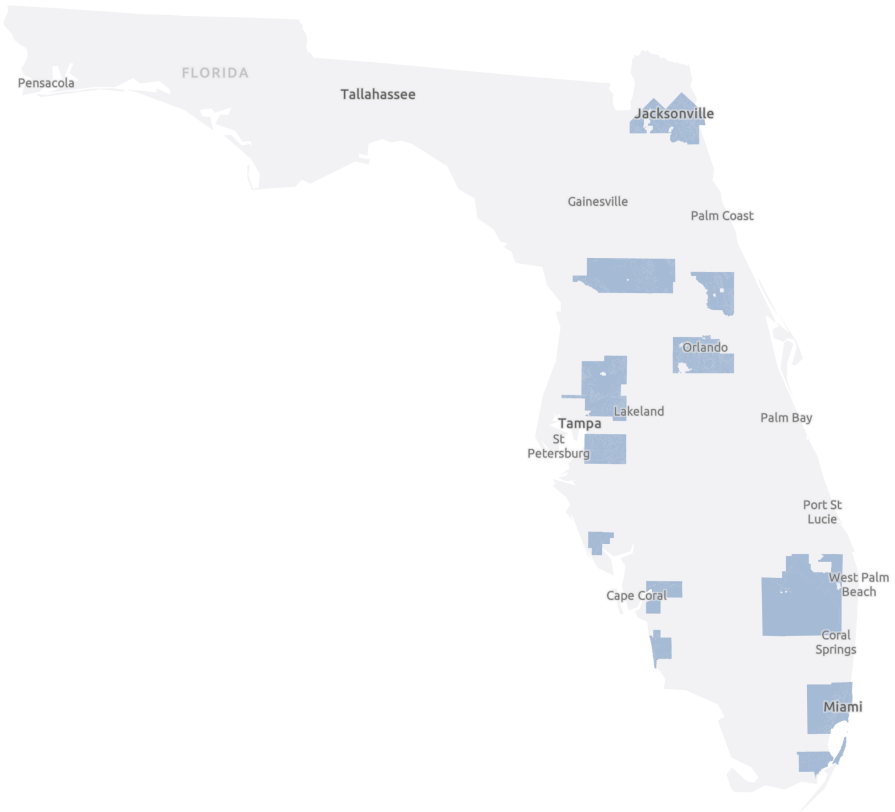
Tables and Figures

Figure 2-1: Digitized Flood Maps

(a) Digitized Map of Miami (1978)



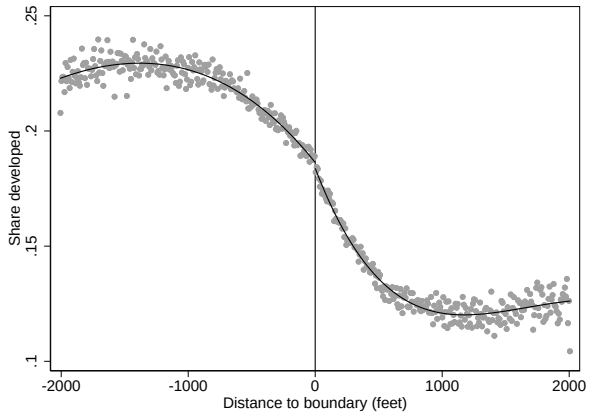
(b) Coverage of Digitized Maps



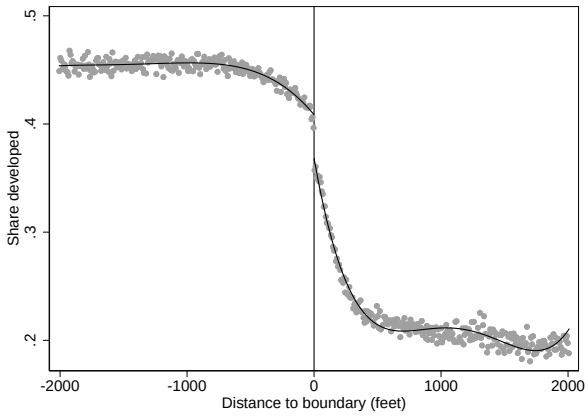
Notes: Figure (a) shows an example of a historical flood map in Miami. flood zones are depicted in blue. Figure (b) displays the coverage of our digitized sample.

Figure 2-2: RD Estimates: Development

(a) Pre-Period Share Developed (about 1980)



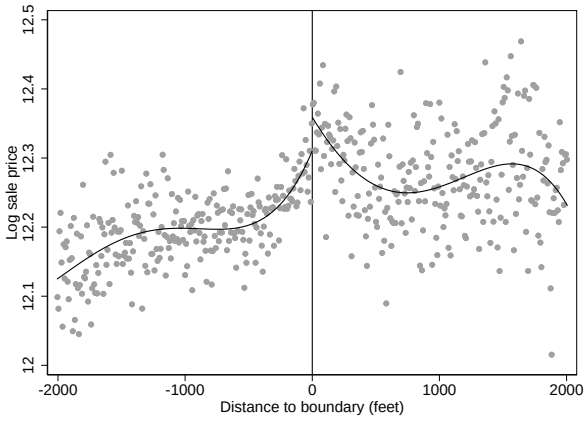
(b) Share Developed in 2016



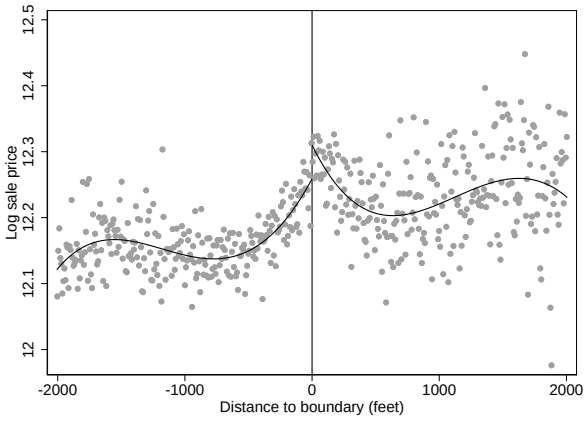
Notes: Figures present RD plots with a fourth order polynomial fit on either side of the flood zone boundary. Distance to boundary is measured in feet, with positive distance indicating being inside the flood zone. Sub-figure (a) plots the share of land that was developed as of the late 1970s and early 1980s, and sub-figure (b) plots the share of land developed as of 2016. Estimates are residualized of census tract fixed effects.

Figure 2-3: RD Estimates: Prices

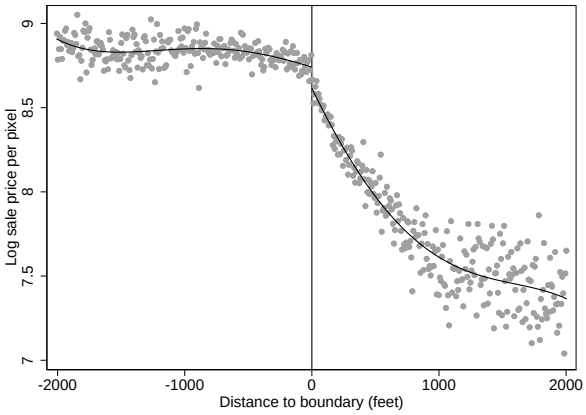
(a) Log Sale Price (2005-2020)



(b) Log Single Family Sale Price (2005-2020)

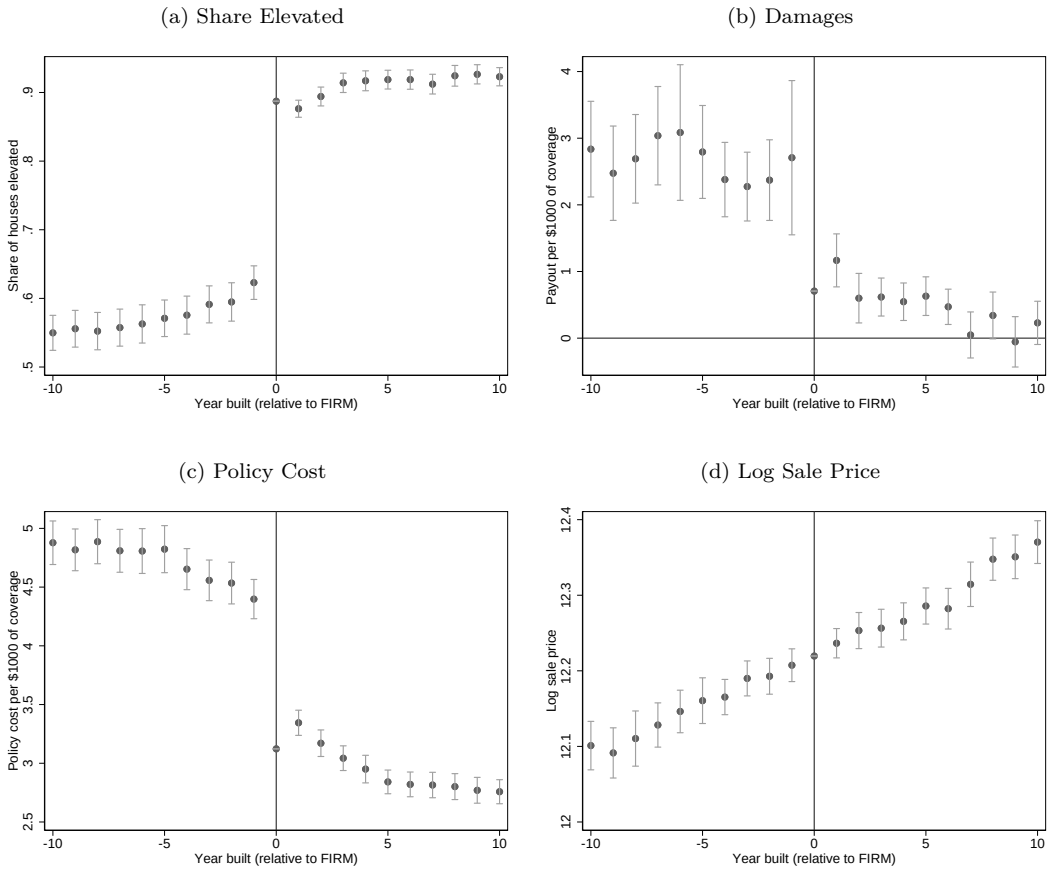


(c) Log Vacant Land Sale Price (2005-2020)



Notes: Figures present RD plots with a fourth order polynomial fit on either side of the flood zone boundary. Distance to boundary is measured in feet, with positive distance indicating being inside the flood zone. Sub-figure (a) plots log sales prices of arms-length sales for homes with structures that sold between 2005 and 2020, sub-figure (b) replicates (a) restricting to single family homes, and subfigure (c) plots sales prices for vacant land, with prices normalized by the size of the lot (results are presented as the sale price per 30x30m pixel). Estimates are residualized of census tract fixed effects.

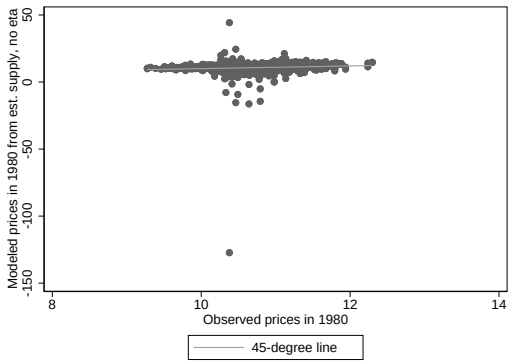
Figure 2-4: NFIP Enrollment Year Event Study Estimates



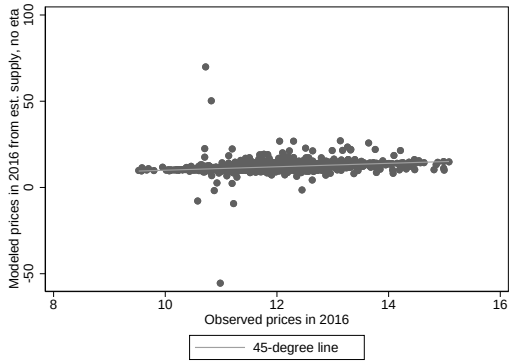
Notes: Figures present coefficients on bins of year built relative to the FIRM year (the year of enrollment in the National Flood Insurance Program, at which time building codes began to be imposed on newly-constructed housing). Sample is restricted to single-family residences inside flood zone flood zones. Sub-figure (a) shows share of NFIP policies that indicate elevation above the modern-day Base Flood Elevation. Sub-figure (b) shows insurance payouts from 2010 to 2018 in each census tract as a share of total dollars of coverage across the time period in that census tract. Sub-figure (c) shows total policy costs to consumers from 2010 to 2018 in each census tract as a share of total dollars of coverage across the time period in that census tract. Sub-figure (d) shows the log sale price in 2010 dollars between 2005 and 2020 in each census tract.

Figure 2-5: Supply Model Reliance on Structural Error Terms

(a) Prices to rationalize 1980 development using supply curve, omitting idiosyncratic costs

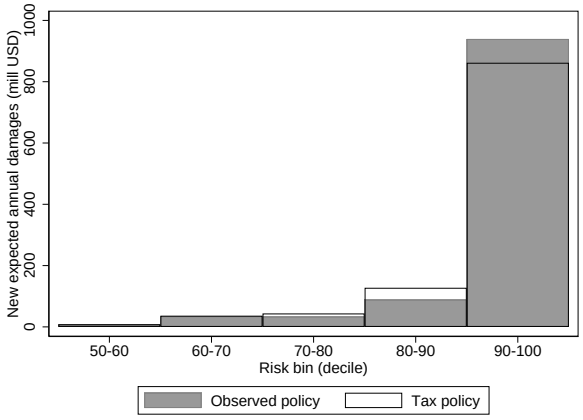


(b) Prices to rationalize 2016 development using supply curve, omitting idiosyncratic costs



Notes: Figures present prices that rationalize observed development using just the estimated supply curve, without estimated idiosyncratic construction costs. Sub-figures (a) and (b) present the prices in 1980 and 2016 that would rationalize the observed quantities of development, using the estimated parameters (μ_j, σ_j, ψ) and observed elevation decisions, if the model omitted structural supply costs.

Figure 2-6: Expected new annual damages by risk bin: status quo vs. first-best corrective tax policy



Notes: Figure plots expected annual damages on newly-constructed houses by risk bin under the observed and counterfactual tax policies. Sample includes both flood zone and non-flood-zone land.

Table 2.1: Summary Statistics of the Spatial RD Sample

	Florida	Digitized map sample		Boundary sample	
		Outside historic flood zone	Inside historic flood zone	Outside historic flood zone	Inside historic flood zone
	(1)	(2)	(3)	(4)	(5)
Panel A. Development					
Share developed in 1980	0.056	0.116	0.124	0.243	0.178
Share developed in 2016	0.145	0.313	0.243	0.473	0.313
Single family homes	5,175,979	552,230	191,507	159,607	80,218
Single family share of structures	0.662	0.872	0.869	0.861	0.869
Share post-FIRM	0.808	0.777	0.648	0.676	0.618
Panel B. Other characteristics					
Share wetlands	0.348	0.205	0.449	0.101	0.343
Share water	0.069	0.015	0.163	0.011	0.098
Distance to coast (miles)	10.7	7.9	6.4	7.7	7.9
Panel C. Prices					
Median house price 1980	47,459	48,198		46,912	
Median house price (2005-2020)	167,233	177,673	274,065	188,148	252,935
Median single family house price (2005-2020)	182,452	179,419	298,716	190,204	277,492
Panel D. Risk					
FEMA flood maps					
Land share in flood zone as of 1996	0.379	0.031	0.798	0.046	0.723
Land share in flood zone as of 2017	0.450	0.106	0.720	0.130	0.664
First Street risk measures					
Land share with $\geq 1\%$ chance of flooding	0.447	0.425	0.582	0.240	0.573
Land share with substantial flood risk	0.153	0.065	0.225	0.061	0.223
Total area (square miles)	58,257	4,169	1,978	746	553

Notes: Table displays summary statistics for the entire state of Florida, the geographic area covered by the digitized flood maps, and a sample restricted to 2,000 feet on either side of the flood zone boundary. Median house price in 1980 is a population-weighted census tract average of 1980 Census estimates of the average value of owner-occupied single family housing, tabulated in 1980 dollars. Median house price (2005-2020) tabulates the median sales price, in 2010 dollars, of houses sold between 2005 and 2020. For houses sold multiple times, we take the average transaction price across sales. House prices from 2005 to 2020 are derived from administrative sales records from the state of Florida and are restricted to arms length sales. Substantial flood risk is defined as areas with an estimated flood depth above two feet.

Table 2.2: Regression Discontinuity Estimates

			Local linear regression		
	Outside flood zone mean	Polynomial	Rectangular kernel, constant bandwidth	Triangular kernel, optimal bandwidth	Polynomial excl. coastal areas
	(1)	(2)	(3)	(4)	(5)
Panel A. Current flood zone status					
In flood zone as of 1996	0.060	0.430	0.405	0.268	0.454
		(0.020)	(0.019)	(0.014)	(0.024)
In flood zone as of 2017	0.130	0.292	0.283	0.207	0.300
		(0.019)	(0.017)	(0.013)	(0.023)
Panel B. Historical land use					
Share of land developed in 1980	0.235	-0.003	-0.005	-0.006	-0.003
		(0.002)	(0.002)	(0.002)	(0.002)
Panel C. Modern land use					
Share of land developed in 2016	0.470	-0.042	-0.044	-0.037	-0.040
		(0.005)	(0.005)	(0.004)	(0.006)
Share of land covered by a building footprint	0.264	-0.029	-0.026	-0.025	-0.027
		(0.004)	(0.004)	(0.004)	(0.004)
Share of land covered by a single family home	0.101	-0.015	-0.013	-0.010	-0.013
		(0.003)	(0.002)	(0.003)	(0.003)
Share of land covered by wetlands	0.098	0.110	0.096	0.071	0.113
		(0.007)	(0.007)	(0.006)	(0.009)
Panel D. Prices					
Log house price	12.151	0.065	0.063	0.066	0.049
		(0.022)	(0.021)	(0.019)	(0.026)
Log single-family house price	12.073	0.056	0.060	0.064	0.051
		(0.018)	(0.018)	(0.015)	(0.021)
Log vacant land price (per 30X30m pixel)	8.806	-0.102	-0.062	-0.070	-0.126
		(0.048)	(0.032)	(0.030)	(0.056)

Notes: Table displays estimates of equation (1). Outside of flood zone means are calculated within 50 feet of the boundary. The polynomial specification estimates a fourth order polynomial separately on either side of the boundary, restricted to a window of 2,000 feet on either side of the boundary. Column (3) estimates linear regressions separately on either side of the cutoff, with each point equally weighted within 250 feet of the boundary. Column (4) estimates the MSE-optimal RD bandwidth from Calonico et al. (2014) and fits a local linear regression within that bandwidth using a triangular kernel. Column (5) replicates Column (2), but excluding land less than one mile from the coast. All discontinuities are estimated on the historic boundaries and exclude boundaries that trace a body of water. Robust standard errors are clustered at the census tract level.

Table 2.3: NFIP Enrollment Regression Coefficients

Variable	(1) Mean in flood zone, Pre-FIRM	(2) Event study	(3) DD Spec
Share elevated	0.57	0.271	0.335
		(0.013)	(0.019)
Insurance payouts (per $1000 of coverage)	$2.94	$-1.61	$-2.84
		(0.40)	(0.96)
Policy cost (per $1000 of coverage)	$4.80	$-1.21	$-1.78
		(0.075)	(0.127)
Log house price (sold 2005-2020, in 2010 $USD)	12.21	-0.007	0.004
		(0.009)	(0.011)

Notes: Table presents variable means and coefficient estimates on Equation 2.3 from the event-study and difference-in-difference analyses of NFIP enrollment on share elevated, insurance payouts, policy cost, and house price. Elevation, payout, and cost data come from residential NFIP policies from 2010-2018. Price data come from residential sales prices in 2005-2020. Sample includes all single-family residences in Florida. Standard errors are clustered at the census tract level.

Table 2.4: Parameter Estimates

Supply cost of flood zone designation (ψ)	0.246
	(0.296)
Demand cost of flood zone designation ($-\phi$)	0.230
	(0.491)
Demand elasticity (α^D)	-1.0
Consumer WTP to avoid flood zone (ϕ/α^D)	0.230
Estimating α^D?	N

Notes: Table presents estimates of the coefficients on flood zone in the GMM estimation of household preferences and housing supply. Standard errors (in parentheses) were generated from bootstrapping (100 iterations).

Table 2.5: Counterfactual Outcomes

	Modeled Outcomes			Change relative to Unregulated Counterfactual	
	Unregulated	Current Policy	Corrective Tax	Current Policy	Corrective Tax
	(1)	(2)	(3)	(4)	(5)
New constr. on land in observed floodplain					
Approximate N Houses	757,814	640,676	698,942	-117,138	-58,872
				-15.5%	-7.8%
New constr. on all land					
Approximate N Houses		4,472,641		0%	0%
Per-house PDV damages for new construction					
Location-based	$6,245	$5,207	$4,865	$-1,037	$-1,379
(i.e. obs. elevation, counterf. locations)				-16.6%	-22.1%
Elevation-based	$10,770	$5,207	$5,394	$-5,563	$-5,376
(i.e. obs. locations, counterf. elevation)				-51.7%	-49.9%
All	$13,064	$5,207	$5,085	$-7,857	$-7,979
(i.e. counterf. locations & elevation)				-60.1%	-61.1%
Number of Elevated Houses	288,002	640,676	482,901	352,674	194,899
				122%	68%
Price					
Inside observed floodplain	$260,578	$250,865	$251,596	$-9,713	$-8,982
				-3.7%	-3.4%
Outside observed floodplain	$166,954	$174,505	$170,170	$7,550	$3,216
				4.5%	1.9%
Overall	$182,235	$185,670	$182,765	$3,435	$530
				1.9%	0.3%

Notes: Table presents estimates of counterfactual outcomes using the baseline parameters. N houses is defined by assuming that each developed pixel is equivalent to one house. "Observed floodplain" is the area designated as the flood zone in the original flood maps. Areas in our sample counties without digitized historic flood maps are assigned their flood zone status as of 1996, which we estimate overlaps with historic flood zone status in over 90% of cases. Prices are weighted by total developed area. The unregulated counterfactual sets $SFHA_z = 0$ everywhere. The corrective tax counterfactual sets $SFHA_z = 0$ everywhere and imposes taxes equal to expected flood damages, conditional on socially-optimal elevation choices.

Table 2.6: Counterfactuals: Welfare-Relevant Components

Outcome (Millions of $)	Level	Differences from Unregulated		
	(1) Unregulated	(2) Current Policy, $\phi_1 = \phi$	(3) Current Policy, $\phi_1 = 0$	(4) Corrective Tax
Producer Surplus		-3,095	-3,095	-2,188
Consumer Surplus		-62,377	-39,192	-41,462
Damages	58,431	-35,140	-35,140	-35,687
Government Revenue	3,225	8,524	8,524	20,117
Total Social Welfare		-21,809	1,377	12,154

Notes: Table presents estimates of counterfactual outcomes using the baseline parameters. Outcomes are in millions of $. The unregulated counterfactual sets $SFHA_z = 0$ everywhere. Column (2) computes consumer surplus under the current policy, assuming that the entire flood zone demand cost affects experienced utility. Column (3) replicates column (2), but allows most of the flood zone demand cost to reflect debiasing rather than experienced utility costs. Column (4) describes outcomes in the corrective tax counterfactual, which sets $SFHA_z = 0$ everywhere and imposes taxes equal to expected flood damages, conditional on socially-optimal elevation choices. Government revenue indicates revenue from insurance premiums in columns 1-3, and revenue from the corrective tax in column 4.

Chapter 3

Homeowner, Mortgage Lender, and Homebuyer Responses to Homeowner's Insurance Prices

By Abigail Ostriker and Anna Russo[1]

Abstract

Climate change is anticipated to substantially increase property insurance premiums, reflecting heightened physical risk. We investigate whether and how property insurance prices affect real estate markets in the state of California. We leverage cross-ZIP code variation in homeowners' insurance prices that is driven by differential insurer exposure to wildfires elsewhere in the state. We find that high exposure to insurer losses in other markets increases premiums by more than 3% per year and only marginally decreases insurance coverage. Higher premiums lower the share of house purchases with a mortgage, indicating that bank unwillingness-to-lend dominates any increased desire of homebuyers to shift their risk to banks. Higher premiums also elevate house sales, demonstrating that existing homeowners' expanded willingness-to-sell dominates any erosion of demand from prospective buyers. Finally, higher premiums have no detectable effect on house prices, showing that premium increases of this magnitude are unlikely to trigger a housing market collapse.

3.1 Introduction

The acceleration of climate change, with increasingly frequent and destructive natural disasters, is spurring fears that property insurance markets may collapse (Tavernise and Flavelle, 2022). In the wake of record-setting wildfires and hurricanes, major property insurers have begun pulling out of high-risk areas (Flavelle,

[1]Ostriker: Department of Economics, MIT, ostriker@mit.edu; Russo: Department of Economics, MIT, aerusso@mit.edu.
Ostriker gratefully acknowledges support from the Hausman Dissertation Fellowship.

2019; Rouse et al., 2023). In addition to limiting individuals' ability to smooth consumption, reduced insurance availability could cool real estate markets. The greater expense and perhaps higher anticipated risk of owning a house with expensive insurance premiums may diminish house prices. Furthermore, because mortgage lenders require borrowers to maintain homeowner's insurance, if policies become prohibitively expensive or if no insurer is willing to issue a policy, it could prevent hopeful buyers from obtaining mortgages (Flavelle, 2022; Keys, 2023). Because few buyers have the means to purchase in cash, this restriction in mortgage availability would further reduce demand and house prices. A collapsing real estate market could then, some worry, put the "health and survival" of entire state economies at risk (Tavernise and Flavelle, 2022).

Despite these fears, little is known about how real estate markets respond to property insurance prices. While higher premiums could entail greater expense and signal larger risk, they may be difficult to observe while shopping for a home, so demand for these houses may be largely unaffected. Mortgage lenders may be similarly unresponsive to premiums, or they could take rising premiums as a signal of higher risk and reduce their willingness to lend. In addition, real estate prices also depend on homeowners' inclination to sell their homes, which may be accentuated by higher insurance premiums. To understand the importance of these channels, we study the impact of homeowners' insurance prices on housing transactions in the context of California's real estate market.

In recent years, California experienced multiple devastating and record-setting wildfires across different parts of the state. Insurers "reacted swiftly" to this "wake-up call", with many companies attempting to increase rates (Xu et al., 2019). However, because of California's restrictive regulations, only insurers who actually experienced losses were able to raise their premiums. Conditional on risk levels, wildfire losses are randomly allocated across insurers by the stochastic nature of the disaster. Because insurer market shares vary across ZIP codes, some areas are more exposed to these losses than others. From the perspective of a given ZIP code, price increases driven by random exposure to wildfire losses are arguably exogenous to other market-level trends. We therefore leverage this quasi-experimental variation in insurance prices in a difference-in-difference design to study the impact of insurance availability on the real estate market.

Our analysis compares ZIP codes that were more- versus less-exposed to insurer losses in 2017, the first year of extreme wildfires. We restrict to ZIP codes which were not themselves exposed to wildfires in 2017. We combine data on homeowner's insurance prices and policy counts, obtained via a public records act request, with counts of mortgage originations, which are made publicly-available under the Home Mortgage Disclosure Act. We also incorporate data on housing sales to measure the quantity and price of transactions. For all outcomes, ZIP codes which are more exposed to insurer losses exhibit parallel trends to the rest of the state in the years preceding the wildfires. However, after 2017, insurance premiums in the most-exposed group rise by more than 3% per year. The number of purchased policies only marginally declines, showing that the premium increases do not substantially reduce demand for insurance. Although insurance is still available to homebuyers, the share of house purchases with a mortgage falls by about 5%, indicating that higher insurance premiums restrict the supply of mortgages. Despite this reduction in mortgage availability, house sales accelerate in treated ZIP codes, with houses being 6% more likely to be sold in 2020 than in 2017. This increase in quantity of house sales demonstrates the dominance of expanded housing supply over weakened housing demand. Nevertheless, we do not find evidence of a reduction in sales prices, indicating that insurance premium hikes of this magnitude are unlikely to impact property taxes or spill over to the rest of the economy.

This paper relates to a few strands of the literature. Firstly, we contribute to a body of work studying factors that impact demand for property insurance. Several papers study how factors such as risk salience (Gallagher, 2014; Royal and Walls, 2019; Mulder, 2022), home equity (Liao and Mulder, 2021), and risk levels (Bradt et al., 2021) impact demand for property insurance. We supplement these analyses by studying the role of prices, finding a low elasticity of demand. Secondly, this paper also contributes to a growing literature on mortgage lender responses to natural disaster risk. Several recent papers find that mortgage lenders are reducing their lending in risky areas (Keys and Mulder, 2020; Sastry, 2021; Blickle and Santos, 2022; Ouazad and Kahn, 2022). Moreover, lenders in the United States are also shifting riskier loans to the federal government (Keenan and Bradt, 2020; Ouazad and Kahn, 2022). Relative to these papers, we study mortgage lender responses to insurance prices, not to environmental risk itself. Studying the impact of homeowner's insurance specifically is valuable because it is one of the main tools with which the housing market can mitigate the negative consequences of climate change. Finally, we contribute to a literature examining the capitalization of location-specific fees into housing prices (Palmon and Smith, 1998a,b; Gallagher et al., 2013; Livy, 2018). Relative to these studies, which generally find property taxes are close to fully-capitalized into home prices, we find limited capitalization in the context of insurance premiums. Our results suggest that insurance prices substantially reduce existing homeowners' willingness to stay in their houses, as well as decrease willingness of banks to issue mortgages, so areas with increasing levels of climate risk could witness an exodus in the future.

3.2 Conceptual Framework

An increase in homeowner's insurance premiums could impact the housing market through several different channels. On the homeowner's (supply) side, higher premiums could increase the expected cost of staying; homeowners might then seek to sell their house in order to move to a less-expensive location. The expected cost of staying could increase both directly, from the premium increases, and indirectly, if homeowners take insurance prices as a signal of uninsured risk. Homeowners may not purchase complete coverage, leaving them exposed to damage that exceeds the coverage cap. Also, homeowners cannot insure against the cost of losing sentimental possessions, the inconvenience of relocating for several months as a house is repaired, or the trauma of evacuating under the threat of an imminent wildfire. Without any change in demand for housing, this expansion of supply would be expected to increase the quantity and decrease the price of houses sold.

However, higher homeowner's insurance premiums may also impact demand for housing. Potential homebuyers, if they are aware of the premiums, might be less interested in buying expensive-to-insure houses for the same reasons that homeowners want to sell. Lower demand for these houses would diminish the quantity of houses sold as well as their price. Furthermore, even if buyers still wanted to purchase, banks might be less willing to issue a mortgage for a more-expensive-to-insure house. While banks typically require borrowers to maintain homeowner's insurance for the duration of the mortgage, homeowners with damage exceeding the coverage cap may choose to default rather than rebuild. If banks take insurance prices as a signal of risk, they might update their estimates of these houses' risk and charge higher interest rates or limit quantity of mortgages they issue. Reduced availability of mortgages could limit the ability of prospective homebuyers to purchase these houses, further reducing demand. Alternatively, if banks evaluate risk using other (non-insurance) signals, insurance premium changes might not affect their assessment. Homebuyers, on the other hand, may have more limited sources of information about risk. If higher insurance premiums cause them to anticipate

higher risk, they could have greater demand for a mortgage because they hope to shift the risk to their lender. The effect of homeowner's insurance prices on the share of purchases with a mortgage is therefore ambiguous.

Combining the supply and demand channels, higher homeowner's insurance premiums are expected to reduce house prices but have an ambiguous impact on quantity transacted. They also could either increase or decrease the share of houses that are purchased with a mortgage rather than cash. Therefore, in this analysis we investigate the magnitude of the house price reduction and both the direction and magnitude of the effects on quantity of transactions and share of transactions with a mortgage. Because homeowner's insurance purchases may be elastic to price, and some consumers may buy homeowner's insurance only to satisfy lender requirements, we also study the impact of insurance prices on share of houses that are insured.

3.3 Context and Data

3.3.1 Property Insurance

A property insurance contract allows a property owner to protect herself against a range of risks, including both non-climate (e.g. vandalism and theft) and climate (e.g. wildfires and wind storms) risks.[2] Homeowner's insurance is the second-largest and fastest-growing type of property insurance in the US (Oh et al., 2021). Banks require mortgage borrowers to purchase homeowner's insurance as a condition of the loan, so coverage is high: up to 80% of homeowners have a policy (Oh et al., 2021).

Our empirical context is the state of California, one of the most populous and wildfire-prone states in the United States. As of 2019, 6.6% of the state's 9.5 million residences were at high or extreme risk of wildfire (CoreLogic, 2019). While all states have some form of insurance rate regulation in order to curb monopolistic practices, California is one of the strictest (Oh et al., 2021). Insurers may not increase rates by more than 7% without triggering a lengthy and expensive hearing process that could span multiple years (American Property Casualty Insurance Association, 2020). Furthermore, when filing for rate increases, insurers may not base rates on "modern catastrophe models" that incorporate scientific and engineering assessments of risks, but are required to use historic catastrophe losses instead (American Property Casualty Insurance Association, 2020). These restrictions imply that insurers who are randomly exposed to higher wildfire losses are able to successfully argue for higher insurance prices. This phenomenon will yield quasi-experimental variation in rate increases, forming the basis of our research design.

3.3.2 Data

In addition to providing quasi-experimental variation, the setting of California also provides uniquely useful data to study the impacts of insurance availability. We describe these data, and other data we use for this analysis, below.

[2]Floods and earthquakes are not typically covered risks.

Insurance Policies To measure prices and counts of policies, we use data on the dollar amount of premiums earned and exposures earned, by insurance company, ZIP code, and year for home and fire insurance policies written in 2015-2020. These data were obtained from the California Department of Insurance (CDI) and were collected as part of their Community Service Statement survey. The policies described include policies in a number of lines of business, including but not limited to homeowner's policies and dwelling-fire owner-occupied policies.[3] Few other states make this type of data available to the public.

Insurance Losses We also make use of insurance loss data reported in the Annual Reports submitted by insurers to the state of California. These data report, for each year, total written premiums, losses, market share, and loss ratios of each insurer (NAIC code). These data were also obtained via a public records act request. We focus on losses in 2017, the first year of record-setting wildfires in the state.

Wildfire Perimeters We define ZIP codes as being exposed to a wildfire if it intersects anywhere with a wildfire perimeter from the California Fire and Resource Assessment Program (FRAP) dataset. We drop ZIP codes that were exposed to a wildfire in 2017.

Mortgage Originations To measure mortgage originations, we use data collected by the Consumer Financial Protection Bureau under the Home Mortgage Disclosure Act (HMDA) from 2015-2021. These data describe characteristics of new mortgages, at the loan level, geocoded to the Census tract level. To isolate new home purchases, we restrict to loans used to purchase a home that are secured by a first lien. We also restrict to single-family buildings that are owner-occupied as a principal dwelling and drop manufactured homes, in order to better align with our denominator of single-family house sales. We measure the total number of originated mortgages. We also incorporate tract-level data on the population and number of housing units compiled by the US Census and appended to the public HMDA data. We merge the tract-level data to ZIP Code Tabulation Areas (ZCTAs) based on the share of a tract's housing units that are in each ZCTA, using the crosswalk provided by NHGIS (Manson et al., 2021).

House Transactions We acquire additional metrics regarding housing transactions from Redfin, a national real estate brokerage.[4] Redfin compiles these data from Multiple Listing Service entries for each ZIP code in the US. We restrict to single-family homes for years 2015-2020. We measure number of home sales as the total number of homes with a sale date in a given year, and the median sale price per square foot as the median final home sale price divided by the total square footage of the built property (not the lot) in each year. We calculate an implied average square footage as the ratio of median sale price to median price per square feet, and drop observations if the implied mean square footage falls below 100.

Wildfire Risk Finally, we bring in parcel-level data describing wildfire risk from the First Street Foundation, a nonprofit devoted to quantifying and communicating climate risks using state-of-the-art modeling techniques. This dataset provides, at the parcel level, a "Fire Factor" risk score from 1-10, with 10 being the highest

[3]Among insurers who wrote at least $5 million in premiums for either Dwelling Fire or Homeowners lines of business, these policies made up about 55% of the total number of issued policies, according to data from the Personal Property Experience data call also provided by the CDI.

[4]https://www.redfin.com/news/data-center/

wildfire risk and 1 being the lowest. It also provides a variable describing the modeled annual probability of burning in a wildfire. We aggregate these measures to the ZIP code level.

After aggregating our variables of interest to the ZIP code level, we construct our variables of interest before creating a balanced panel from 2015-2020.

3.3.3 Variable Definitions

We define premiums at the insurer level and ZIP code level as the total dollar amount of earned premiums divided by the total number of earned exposures.[5] In order to account for differently-sized ZIP codes, we normalize number of policies, total home sales, and number of mortgage originations by the number of owner-occupied housing units in each ZIP code. Because insurance policies can also be taken out for non-owner-occupied units, this is an underestimate of the total size of the insurance market. However, we obtain similar results if we instead normalize by number of 1-4 family housing units.

Our empirical strategy leverages variation across ZIP codes in the statewide losses borne by insurers in those ZIP codes. We capture this variation by defining a ZIP-level measure of insurer average losses in 2017. For each ZIP code, we define the weighted loss ratio in 2017 as the average of insurer loss ratios in 2017, weighted by insurer market shares in 2017.[6] That is, after excluding insurers with missing information on loss ratios in 2017, we define the ZIP-level average loss ratio in 2017 as:

$$LR_z = \frac{\sum_{j \in z} LR_j N_{jz}}{\sum_{j \in z} N_{jz}} \tag{3.1}$$

In order to study the decision to purchase a house in cash versus with a mortgage, we create a variable capturing mortgages per house sale. We define the mortgage-per-sale ratio as our measure of single-family, first-lien, owner-occupied mortgage originations divided by sales of single-family houses.

Finally, we drop ZIP codes with a mortgage-to-sales ratio exceeding 10 in any year, due to concerns that Redfin may have underreported sales in some ZIP-years.

3.3.4 Summary Statistics

Table 3.1 presents summary statistics, both across all ZIP codes and separately for the top and bottom 3 quartiles of average loss ratio. Observations are at the ZIP-year level, restricting to the pre-period (i.e. 2015-2017). As might be anticipated, the top quartile differs from the rest in a number of dimensions. However, the variance is large and most differences are not statistically-significant. Average premiums are about 50% higher in the top quartile compared to the rest: \$1400 per year compared to \$850. However, this is not because of higher wildfire risk: First Street Foundation gives an average risk rating of 2.5 out

[5] In contrast to "written" exposures, which is the total amount of coverage associated with a policy at the time of its writing, "earned" exposure at a certain point in time reflects the share of written exposures for which coverage has already occurred. Earned exposures therefore capture realized insurance coverage rather than future coverage.

[6] The loss ratio is an insurer's losses divided by earned premiums in a given year.

of 10 in top-quartile ZIP codes, compared to 2.7 out of 10 in remainder ZIP codes. Similarly, the annual burn probability is 4 hundredths of 1 percent among top-quartile ZIP codes, compared to 8 hundredths of 1 percent in the remaining group. Instead, this difference is likely due to higher home values: the average house sale price in the top quartile is almost $1.2 million, nearly twice as much as the average in the rest of the distribution. This difference is partially driven by higher per-square-footage prices, which are about 50% higher in the top quartile. Incomes in the top quartile are also higher, about $76,000 per year, compared to $68,000 per year in the remaining ZIP codes. In other dimensions, the ZIP codes appear similar: Top-quartile and rest-of-distribution ZIP codes are about equally likely to involve a mortgage in a home purchase. They also have similar rates of wildfire exposure, though the top quartile is slightly more likely to experience a wildfire in later years.

Histograms of our outcomes of interest are presented in Appendix Figure B-1.

Table 3.1: ZIP Code Summary Statistics: 2015-2017

	All ZIPs		Top Quartile Avg Loss Ratio in 2017		Bottom 3 Quartile Avg Loss Ratio in 2017	
	Mean	SE	Mean	SE	Mean	SE
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Insurance Data						
ZIP Loss Ratio in 2017	1.9	0.3	2.3	0.2	1.8	0.2
Premium Price	$975	$508	$1,417	$818	$865	$308
N Policies per year	5,397	4,045	4,013	3,310	5,743	4,138
N Policies per owner-occupied unit	1.59	17.04	1.36	2.61	1.65	19.01
Panel B. Housing Market Outcomes						
N Single-family house sales	232.7	216.4	171.3	167.0	248.0	224.5
N Sales per owner-occupied unit	0.060	0.296	0.096	0.650	0.051	0.060
N Mortgage originations	223.5	208.4	153.4	147.5	241.0	217.6
N Mortgage originations per sale	1.05	0.69	0.98	0.75	1.07	0.67
Median single-family house sale price	$679,273	$611,160	$1,124,994	$925,970	$567,842	$436,722
Median sale price per square foot	$375	$247	$528	$340	$336	$200
Panel C. Fire Risk						
Fire Factor Rating	2.6	1.9	2.5	1.4	2.7	2.0
Annual Parcel-level Burn Probability	0.0007	0.0015	0.0004	0.0008	0.0008	0.0016
Wildfire in 2015	0.042	0.201	0.043	0.203	0.042	0.200
Wildfire in 2016	0.055	0.227	0.067	0.250	0.051	0.221
Wildfire in 2017	0	0	0	0	0	0
Wildfire in 2018	0.075	0.263	0.072	0.258	0.075	0.264
Wildfire in 2019	0.065	0.247	0.086	0.281	0.060	0.237
Wildfire in 2020	0.107	0.309	0.134	0.341	0.101	0.301
Panel D. Demographics						
Population in 2010	27,933	21,135	16,748	14,199	30,730	21,655
N owner-occupied housing units	5,129	3,811	3,992	3,237	5,413	3,890
N 1-4 family housing units	7,669	5,363	5,721	4,598	8,156	5,430
Median annual household income	$72,910	$18,243	$83,418	$22,743	$70,283	$15,886
N ZIP Codes	1,045		209		836	

Notes: Table presents summary statistics, for (1) the entire sample, (2) the treatment group, consisting of ZIP codes in the top 25% of average insurer loss ratio in 2017, and (3) the control group, consisting of ZIP codes in the bottom 75% of the distribution. Observations are at the ZIP-year level, restricted to years 2015-2017.

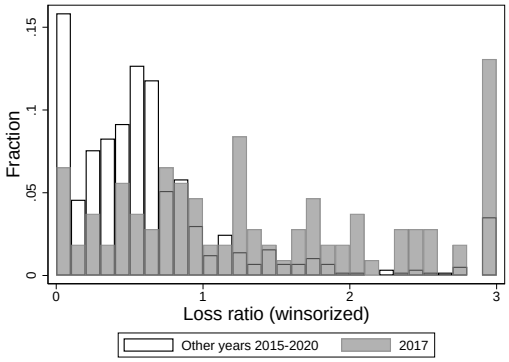
3.4 Empirical Strategy

Our empirical strategy will be based on cross-ZIP code variation in insurer exposure to losses across the entire state of California in 2017, the first year of record-setting wildfires. Figure 3-1 presents histograms of insurer-level and average ZIP-level loss ratios, both in 2017 and other years. As the insurer-level histogram

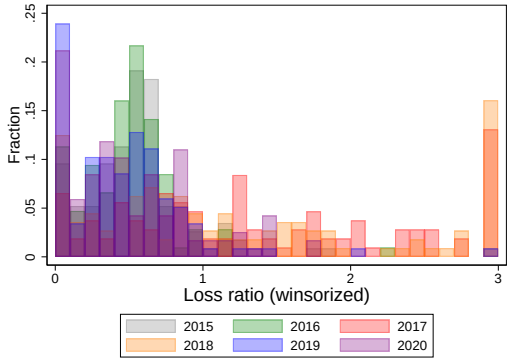
shows, losses in 2017 were extremely high relative to the distribution of losses in other years (except 2018). However, exposure to losses varies widely across insurers. Even when aggregated to the ZIP code level, a substantial amount of variation in loss ratios remains.

Figure 3-1: Distribution of loss ratios

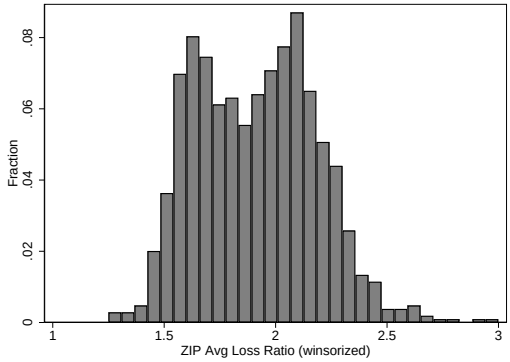
(a) Loss ratio in 2017 vs Other Years (insurer level)



(b) Loss ratio, by Year (insurer level)



(c) Average loss ratio in 2017 (ZIP code level)



Notes: Panel (a) shows a histogram of insurer loss ratios (the ratio of losses to earned premiums) in 2017 compared to other years. Panel (b) breaks out the loss ratios in other years into each specific year. Panel (c) shows a histogram of ZIP code level loss ratios in 2017, where insurer-level losses are averaged within ZIP code, weighting by 2017 market share.

Since insurer losses in 2017 were primarily driven by the random and unpredictable path of wildfires, ZIP code-level exposure to wildfire losses is arguably exogenous to trends in premiums, policy counts, and mortgage originations. We present maps of ZIP-level loss ratios in Appendix Figure B-2 to bolster this argument. The maps show that while insurers with smaller losses tend to be located in the southern part of the state, there is not a strong spatial correlation in loss ratios and high-loss ZIP codes are often found near low-loss ZIP codes. Therefore, ZIP codes with relatively high exposure to losses in 2017 could have been expected to follow a parallel path to less-exposed ZIP codes in the absence of these losses.

In practice, however, because of California's strict regulatory environment, ZIP codes with high exposure to losses experienced increases in insurance premiums following the 2017 wildfires. California requires insurers to set prices based on historical average losses rather than catastrophe simulation models, which enhance

risk predictions by combining location-specific data with scientific models (Xu et al., 2019).[7] Because of this restriction, including losses from the wildfires in 2017 allowed some insurers to substantially increase their premiums.

We leverage this plausibly-exogenous variation in insurance premiums to study the effect of higher premiums on bank willingness-to-lend, prospective buyers' willingness-to-purchase, and homeowners' willingness-to-sell. We compare ZIP codes with the highest exposure to insurer losses to those with less exposure. We examine how more- and less-exposed ZIP codes differ in their average premiums, number of insurance policies purchased, number and price of houses sold, and number of mortgages originated per home sale.

Time Trends To provide preliminary evidence on treatment effects, Figure 3-2 plots trends over time in our outcomes of interest, split by quartile of the ZIP-code-level average loss ratio in 2017. We see that prior to 2017, all outcomes of interest follow parallel trends, which bolsters our identifying assumption. However, after 2017, some differences emerge.

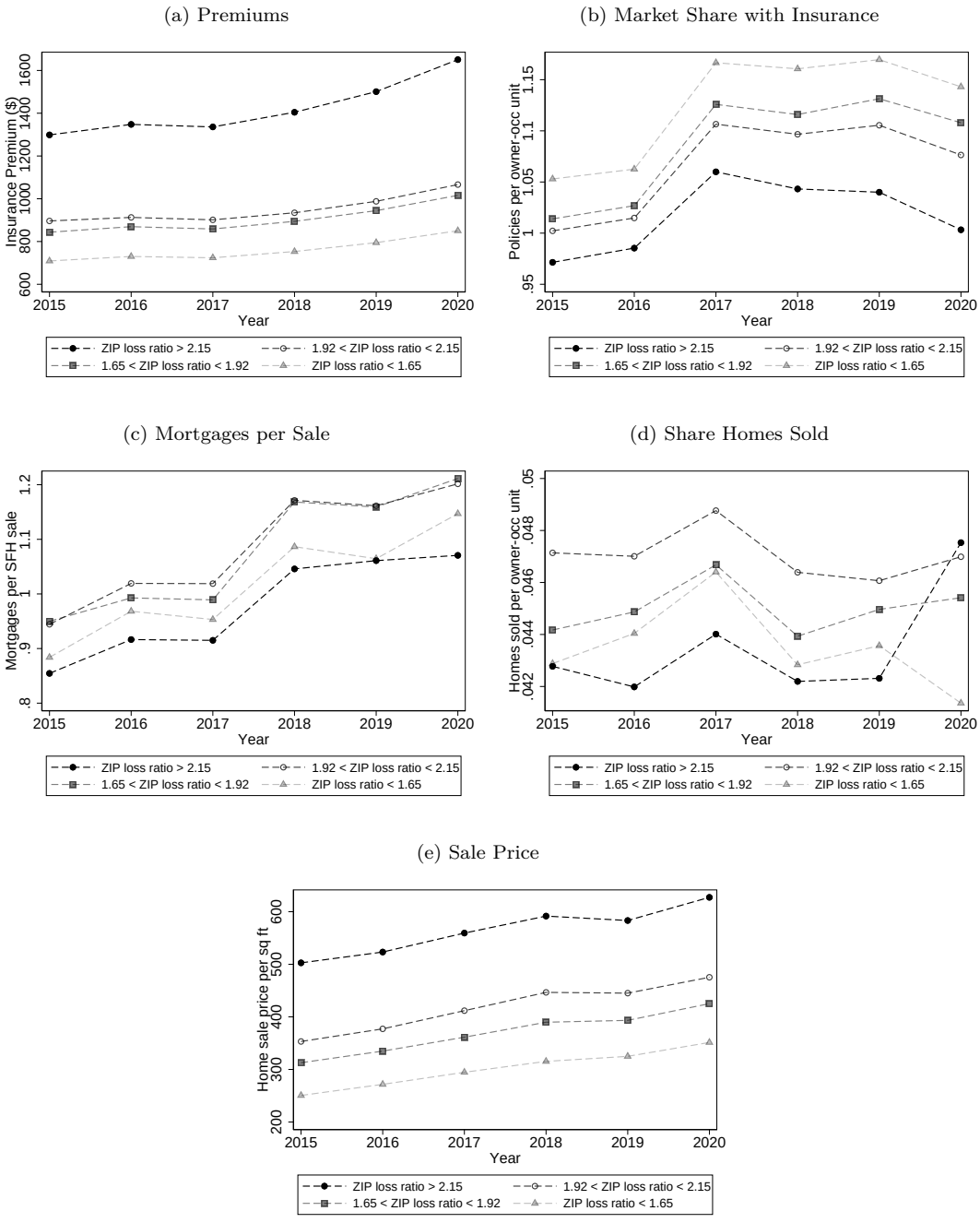
Insurance prices increase after 2017 in all groups, but the growth is particularly steep in the top quartile. In the top quartile, the number of homeowner's insurance policies also begins to diverge from the other ZIP codes after 2017, shedding policies.

The number of mortgages per home sale is lower in the most-treated group than the less-treated groups, but it moves in parallel with them prior to 2017. In subsequent years, share with mortgage is flat in the most-treated group, while it grows in the less-treated quartiles. Similarly, in the post-period the number of homes sold increases the most in the most-treated areas and the least in the least-treated areas. These mirror-image behaviors in the most- and least-treated quartiles support a causal attribution of the changes to the intensity of insurer losses.

Sale prices appear to move in parallel for the duration of the study period. Simple models of the housing market would predict that insurance premium increases would tend to reduce house prices, which could potentially explain the apparent slight dip in prices in the top-quartile ZIP codes in 2019 but does not explain the subsequent return to trend in 2020. However, insurance premium increases on the order of $150 per year might be difficult to detect in these graphs because of their small scale in comparison to the cost of a house.

[7] Earthquake-related perils are an exception to this general rule.

Figure 3-2: Trends over time



Notes: Figure depicts (a) average insurance premium, (b) number of insurance policies per owner-occupied housing unit, (c) number of mortgage originations per single-family housing unit sold, (d) share of owner-occupied houses sold, and (e) median single-family sale price per square foot, by year. Outcomes are at the ZIP code level, split by level of exposure to losses in 2017. ZIP codes are labeled as being in the top (loss ratio exceeds 2.15), second (1.92-2.15), third (1.65-1.92), and bottom (loss ratio falls below 1.65) quarters of the distribution.

3.4.1 Difference-in-Difference Analysis

In this and the following section, we quantify these observed trends and interpret them further. We begin with an event study analysis that compares the most-treated ZIP codes to less-treated ZIP codes before and after the wildfire year (2017). We define the most-treated ZIP codes as those with a weighted average loss exceeding the 75th percentile among our analysis sample. We estimate the following model:

$$y_{zt} = \Sigma_{t\in[-2,3]}\beta_t Treat_z + \delta_t + \zeta_z + \varepsilon_{zt} \tag{3.2}$$

where t denotes year relative to 2017, z denotes ZIP code, and $Treat_z$ indicates a ZIP code with a top-tier average loss. We include time (δ_t) and ZIP code (ζ_z) fixed effects. In an alternative specification, instead of controlling for ZIP code fixed effects, we instead control for the ZIP-level fire risk rating and estimated burn probability, as well as ZIP-by-year indicators of the presence of a wildfire. Our coefficients of interest, β_t, describe how outcomes of interest in treated ZIP codes vary relative to control ZIP codes, before and after the wildfires in 2017. We estimate this model for the outcomes of interest, namely: premiums, number of policies and house sales per owner-occupied housing unit, number of mortgage initiations per house sale, and house sales price per square foot. We weight by number of owner-occupied housing units in the ZIP code.

Figure 3-2 shows that the divergence between more- and less-treated ZIP codes intensifies over time, with the largest differences between top and bottom quartiles appearing in the latest year in the sample. This pattern indicates the impact of the wildfires in 2017 manifests as a slope shift, not a level shift. We therefore summarize the difference-in-difference analysis by estimating a single post-treatment slope shift in the following model with the same outcomes:

$$y_{zt} = \alpha + \beta Y_t Post_t Treat_z + \delta_t + \zeta_z + \varepsilon_{zt} \tag{3.3}$$

where Y_t indicates the year relative to 2017 and $Post_t$ is an indicator which is positive after 2017.

3.4.2 Pooled Analysis

We supplement the simple treatment-control design with an analysis that leverages heterogeneity in treatment intensities across ZIP codes. This analysis is inspired by the patterns in Figure 3-2, in which differences in the post-period appear to increase over time proportionately to the intensity of ZIP code losses. We model the impact of insurer losses as affecting outcomes in proportion to the magnitude of the losses. That is, we estimate the model:

$$y_{zt} = \Sigma_{t\in[-2,3]}\beta_t LR_z + \delta_t + \zeta_z + \varepsilon_{zt} \tag{3.4}$$

where LR_z denotes the ZIP-code-level 2017 loss ratio, normalized by subtracting the mean cross-ZIP loss ratios and dividing by the standard deviation. Our coefficients of interest, β_t, therefore indicate the impact of increasing a ZIP code's average insurer loss ratio by 1 standard deviation. In this design, we have assumed as before that in the absence of treatment, all ZIP code-level outcomes would progress in parallel. We have

further imposed a slightly stronger assumption: that the effect of the loss ratio is linear in its magnitude. Figure 3-2 suggests this assumption may be imperfect: prices among the bottom three quarters appear to move mostly in parallel, while home sales and mortgages diverge more substantially among these groups in the post-period. However, making this assumption allows us to leverage the additional variation contained in what was previously the control group, thus yielding more precise estimates.

Again, we weight by number of owner-occupied housing units in each ZIP code and include an alternative specification with different ZIP-code and ZIP-by-year level controls. Also, as before, to summarize these results, we further simplify the model to estimate a single slope shift as a result of higher insurer losses. That is, we estimate the model:

$$y_{zt} = \beta_t Y_t Post_t LR_z + \delta_t + \zeta_z + \varepsilon_{zt} \tag{3.5}$$

where Y_t indicates the year relative to 2017 and $Post_t$ is an indicator which is positive after 2017.

3.5 Results

3.5.1 Yearly Analysis

The yearly impacts of being in a ZIP code with a top-quartile insurer loss ratio estimated from Equation 3.2 are plotted in Figure 3-3. Because the pooled specification yields similar patterns, we present the coefficient estimates from Equation 3.4 in Appendix Figure B-5 and focus the discussion in this section on the difference-in-difference analysis.

Figure 3-3 confirms the patterns suggested by the descriptive analysis in the preceding section. Treated (top-quartile) ZIP codes follow parallel trends prior to the year of wildfires, and then prices begin to increase steeply. By 2020, three years after the triggering event, prices in treated ZIP codes have grown by almost \$150, or about 11%. At the same time, the insured market share falls due to the higher prices by about 1% per year, though these coefficients are not statistically-significant. We take this to mean that insurance demand is relatively inelastic to price. Importantly, this result implies that any other effects we document primarily reflect the impact of insurance prices, not the quantity of insurance policies purchased.

The share of house sales with a mortgage falls by about 5 percentage points for the most-treated group following the increase in prices.[8] This substantial decline implies that growing insurance prices reduce banks' willingness to lend, to an extent that dominates any increase in homebuyers' desire to shift from a cash to a mortgaged purchase. Banks' low willingness-to-lend against houses with higher insurance premiums could signal that they rely on insurance premiums as a signal of risk. If so, banks could be concerned that in case of damages exceeding the insurance coverage cap, buyers may default rather than repair the house and continue to repay the loan. Alternatively or in addition, banks could worry that higher insurance premiums heighten financial pressures on borrowers, reducing timely repayments. The decrease in banks' inclination

[8]This reduction in mortgages appears less strongly in the pooled specification, with a statistically-significant drop only occurring in 2020.

to lend could manifest either as higher interest rates or as mortgage rationing similar to that documented in recent papers by (Sastry, 2021; Blickle and Santos, 2022). Without data on interest rates, we cannot take a stance on the particular mechanism.

Despite the reduction in mortgage availability, statistically-significant increases in number of home sales appear in 2020, two years after the price increases begin. As a consequence of the price increases, houses are about 6% more likely to be sold in 2020 than they were in 2017. The delayed impact of higher insurance prices may be due to the accelerating nature of the price shock: one year after the wildfires, the average insurance premium in a treatment ZIP code was less than $50 per year higher than in a control ZIP code. By three years after the wildfires, the increase had tripled. In addition, California saw another record-breaking year of wildfires in 2020, which may have increased the salience of insurance prices and accentuated differences between low- and high-priced areas.[9]

The fact that house sales increase despite a reduction in bank willingness to lend points to a large increase in the willingness of homeowners to sell expensively-insured houses. This force presumably also acts to reduce prices. We do not see a statistically-significant decrease in post-period; in fact, if anything there appears to be an increase in prices in the top quartile in 2018.[10] However, the fact that the coefficient estimates in 2015 and 2016 are also statistically-significant and less than 0 implies there may be divergent trends in price through this time period. Prices could be independently rising differentially in more-treated ZIP codes, which could bias our estimates upwards. [11]

Taken together, these results suggest that mortgage lenders and homeowners are both less willing to invest or stay invested in houses facing more-expensive insurance premiums. Their reactions to insurance premium growth dominate any reactions of homebuyers, which could indicate that homebuyers are less aware of the premium hikes.

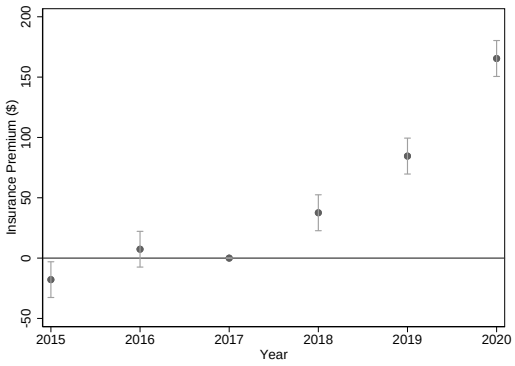
[9] As Table 3.1 indicates, 11% of ZIP codes experienced a wildfire in 2020. Appendix Figure B-3 maps the locations of fires in years other than 2017 in our sample of ZIP codes.

[10] In the pooled specification, where standard errors are smaller, we find that sale prices are substantially higher in more-exposed ZIP codes throughout the post period.

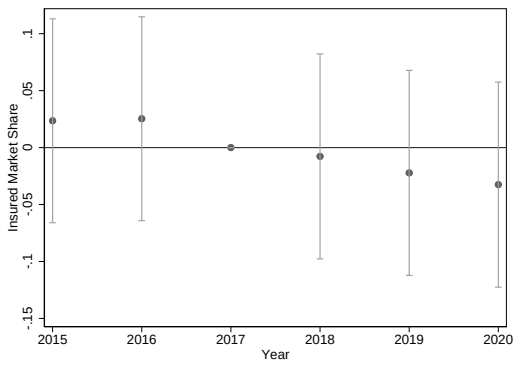
[11] In unpublished results, we add an alternate definition of the treatment group based on the share of a ZIP code's insurance market that was exposed to losses above a certain threshold. Under this specification, the pre-period coefficient confidence intervals contain 0 and the point estimates in 2019 and 2020 fall below 0, though standard errors are large.

Figure 3-3: Difference-in-Difference Annual Coefficients

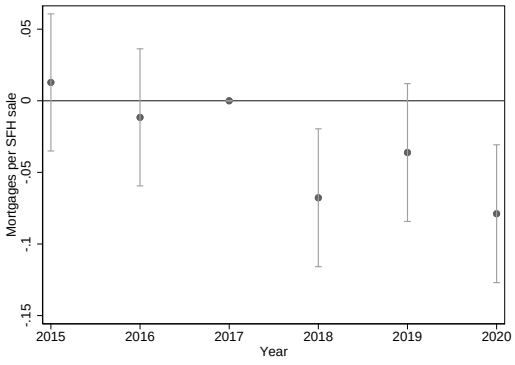
(a) Premiums



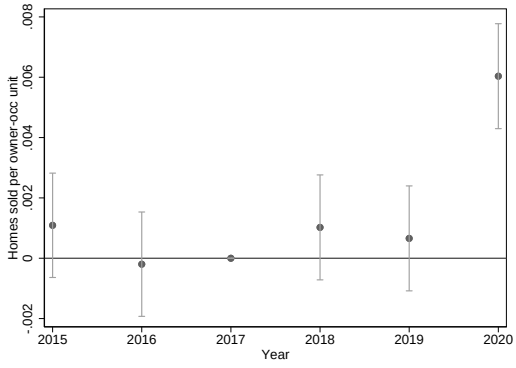
(b) Market Share with Insurance



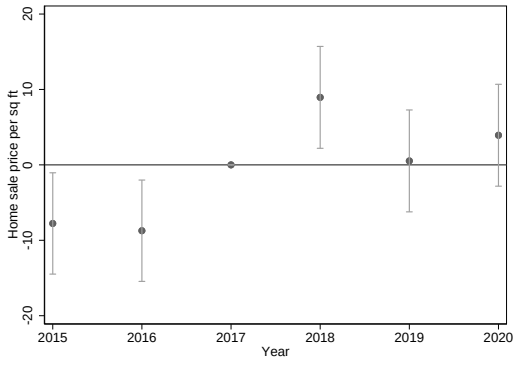
(c) Share of Sales with Mortgage



(d) Share of Homes Sold



(e) Home Sale Price



Notes: Figure depicts estimates of β_t from Equation 3.2 for the outcomes of: (a) average insurance premium, (b) number of insurance policies per owner-occupied housing unit, (c) number of mortgage originations per single-family housing unit sold, (d) share of owner-occupied houses sold, and (e) median single-family sale price per square foot, by year.

3.5.2 Slope-Shift Analysis

To summarize these patterns and evaluate relative effect sizes, we turn to the estimates of the shift in slope induced by higher exposure to insurer losses. We present coefficient estimates for Equations 3.3 and 3.5 in Table 3.2. To aid comparability across specifications, we discuss the magnitudes of the pooled analysis coefficients in terms of the effect of shifting from the 25th to the 75th percentile of insurer average losses, which is an increase of 1.6 standard deviations.

We see that increased exposure to insurer losses substantially increases annual insurance prices. We estimate that being in a top-quartile ZIP code causes premiums to grow by $53, or 3.7%, annually; likewise, shifting from the 25th to the 75th percentile causes premiums to grow by $34, or 3.5%, annually. These estimates are statistically-significant and robust to adding additional controls. We conclude that by 2020, more-treated ZIP codes experience annual insurance premiums that are $100-$150 larger than less-treated ZIP codes.

What impact do these price increases have on insurance take-up? Across all specifications, we see small, negative, non-statistically-significant declines in coverage. In our difference-in-difference specification, we can reject that the top quartile experiences an annual drop in share insured of greater than 3%; in our more-powered pooled specification, we reject that moving from the 25th to 75th percentile causes more than a 0.95% annual drop in share insured. As before, the point estimates are robust to including controls. In the pooled specification, the implied price elasticity of demand is -0.12. The finding that homeowner's insurance takeup is extremely inelastic to price is in line with recent estimates in the context of flood insurance, with Wagner (2022) estimating a price elasticity of demand for flood insurance of -0.25.[12] Homeowner's insurance may be less elastic than flood insurance because banks may more-stringently enforce the requirement that borrowers purchase homeowner's insurance.

Higher insurance premiums also appear to reduce lenders' willingness to issue mortgages, as documented in the previous section. In the difference-in-difference specification, being in the top quartile of insurer losses reduces annual mortgages issued per house sold by about 2.5% per year. This estimate's magnitude, but not its statistical significance, is robust to including controls. The estimated effect appears more muted in the pooled specification: moving from the 25th to the 75th percentile of insurer losses causes the share of home sales with a mortgage to drop by about 1% each year. Again, this is statistically significant only in the specification without additional controls. The marginal statistical significance in the pooled specification is likely due to the effect being concentrated in 2020. Nevertheless, the magnitudes have the potential to be economically significant: if home purchase is at all elastic to whether mortgages are available, a 3% reduction in mortgage availability over a few years could suppress the turnover of houses driven by the increase in supply. It could also shift these expensive-to-insure houses to relatively more wealthy homeowners who can afford to purchase with cash.

In practice, we find that the reduced availability of mortgage does not suppress housing turnover. In fact, we find that higher insurance premiums substantially increase home sales. Both being in a top-quartile ZIP code or moving from the 25th to 75th percentile causes the share of homes sold to increase by 1.7% each year — implying an elasticity of about +0.5. These estimates are marginally statistically-significant in the fully-controlled versions of both specifications, because they are driven by the large effects in 2020. As discussed

[12] Grace et al. (2004) estimate an elasticity of demand around -1 for homeowner's insurance that bundles catastrophe and non-catastrophe coverage. This difference is possibly because their estimation strategy does not account for potential unobserved characteristics of insurance contracts that could be correlated with price.

in the previous section, the increase in quantity of houses transacted reflects the fact that the expansion in homeowner supply outweighs any reductions in demand coming from awareness of higher insurance prices or lower ability to obtain a mortgage. This large increase in willingness-to-sell is particularly striking given the relatively small magnitude of insurance premium increases (up to $150 over three years), compared to the cost of an entire house (more than $1 million in the top quartile). It seems possible that homeowners take large increases in premiums as a signal of risk. Since insurance is incomplete, homeowners may want to avoid that higher risk.

Finally, we would predict that an increase in insurance premiums would be capitalized in to housing prices if buyers and sellers are fully-informed. A persistent increase in annual premiums of $100, taken from our difference-in-difference estimates, would amount to a $2000 cost in present discounted value using a discount rate of 5%. At an average home price of $1.2 million this cost would amount to a reduction of only one-sixth of 1 percent. Phased in over 3 years, this would amount to a price drop of .05% each year. The confidence intervals of our fully-controlled models include this reduction in both specifications, but standard errors are large. The point estimates of the effect of insurance premiums on house prices is actually positive, but as previously discussed, this result could potentially be due to violation of the parallel trends assumption in sales prices. Further work with parcel-level sales prices is needed to ascertain whether an increase in insurance premiums indeed causes sales prices to rise.

Table 3.2: Slope Coefficients

	Difference-in-Difference		Pooled Analysis	
Outcomes	(1)	(2)	(3)	(4)
Panel A. Insurance premium				
$\hat{\beta}$	53.2	53.1	20.6	21.1
(SE)	(1.9)	(13.5)	(0.7)	(5.2)
Mean	1426	1426	978	978
Panel B. Market share insured				
$\hat{\beta}$	-0.017	-0.016	-0.006	-0.005
(SE)	(0.011)	(0.012)	(0.005)	(0.005)
Mean	1.342	1.342	1.895	1.895
Panel C. Share of sales with mortgages				
$\hat{\beta}$	-0.0244	-0.0260	-0.0061	-0.0064
(SE)	(0.0061)	(0.0149)	(0.0024)	(0.0058)
Mean	1.0179	1.0179	1.0845	1.0845
Panel D. Share of homes sold				
$\hat{\beta}$	0.002	0.002	0.001	0.001
(SE)	(0.000)	(0.001)	(0.000)	(0.000)
Mean	0.09	0.09	0.06	0.06
Panel E. Sale prices (per square foot)				
$\hat{\beta}$	3.14	3.23	3.56	3.81
(SE)	(0.86)	(6.35)	(0.33)	(2.39)
Mean	559	559	402	402
Controls				
Year fixed effects	x	x	x	x
ZIP Code fixed effects	x		x	
First street fire factor		x		x
First street burn probability		x		x
Wildfire in ZIP-year fixed effects		x		x

Notes: Table presents estimates of β in Equation 3.3 for the outcomes of average insurance premium, share of market purchasing insurance, number of mortgage initiations per single-family home sale, number of home sales per owner-occupied unit, and sale price per square foot of built property. The mean in columns (1) and (2) represent the mean of the top quartile; the mean in columns (3) and (4) represent all ZIP codes. Columns (1) and (3) are the baseline specification, while columns (2) and (4) replace ZIP code fixed effects with controls for ZIP-code level measures of wildfire risk and include indicators for a wildfire occuring in each ZIP-year.

3.6 Conclusion

We study how real estate markets respond to property insurance price growth in the state of California. We find that areas experiencing higher prices saw minor declines in insurance purchases but substantial increases in house sales, as well as a decline in share of house sales with a mortgage. The magnitude of the shifts relative to the size of the premium increases suggests that both homeowners and banks take higher prices

as a signal of uninsured risk. In response, they expel risky properties from their portfolios. The fact that these expensive-to-insure houses still find buyers could reflect either a lack of information on the part of the buyers, or a reallocation of these houses to a more risk-tolerant population.

Our findings have important implications for policies addressing climate risk in real estate markets. Underpriced insurance can lead to overdevelopment in risky areas (Ostriker and Russo, 2022; Peralta and Scott, 2019) and there is growing policy interest in increasing premiums to counteract this moral hazard (Rouse et al., 2023). Our results shed light on the implications for housing markets of these proposed reforms. However, our conclusions may be limited to areas with relatively low insurance premiums. California's average homeowner's insurance premium is 17% below the national average, and substantially lower than other states facing high environmental risk (American Property Casualty Insurance Association, 2020). In addition, California's house prices are higher than the national average, so insurance premium increases in other states may lead to larger reductions in house prices in percentage terms. Thus, while we find that higher premiums lead to more turnover and limited house price cuts in California, turnover and house prices may both fall in other settings, or if insurance prices increase further.

We leave a few areas open for future work. Although we find that higher insurance premiums do not substantially reduce insurance demand on the extensive margin, coverage could fall on the intensive margin of purchase. Further work is needed to assess responses on this margin. Additionally, our results indicate that mortgages per sale fall when insurance prices rise, suggesting that new buyers may be disproportionately wealthy. It would be interesting to document the characteristics of movers to determine whether the average resilience to property damage grows following a premium increase. Finally, the fundamental determinant of underlying climate risk to property is the quantity of houses built in an area. Investigating the impact of higher insurance prices on new housing construction would be a valuable subject for future study.

Appendix A

Appendix for Chapter 2

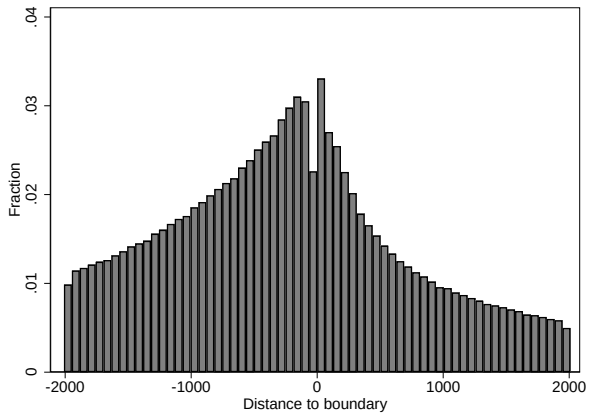
A.1 Tables and Figures

Figure A-1: Building Above the BFE in Naples, FL

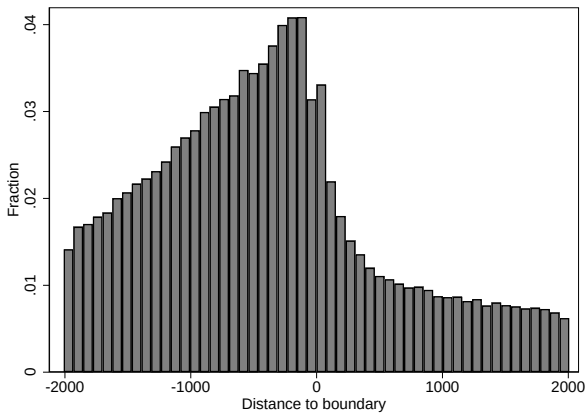
Notes: Figure shows an elevated house in Collier County, Florida. At this location, flood zone regulations require the bottom of the lowest (non-basement) floor to be elevated to 10 feet above sea level.

Figure A-2: Histogram of Distance to Flood Zone Boundary

(a) 30x30m pixels



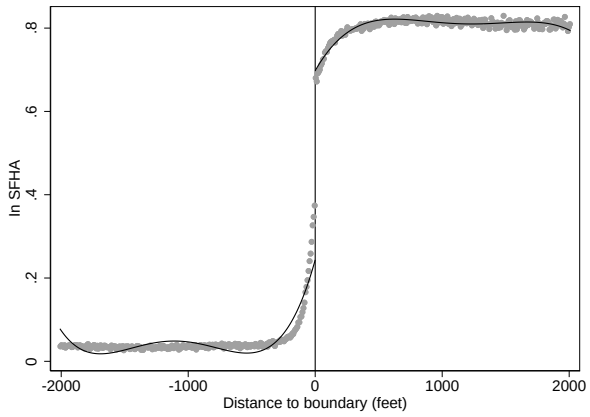
(b) Sales



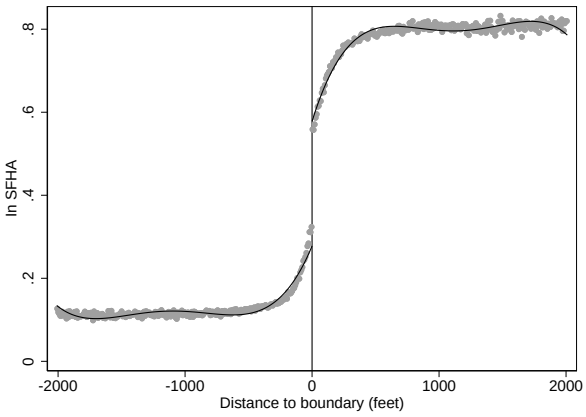
Notes: Figure presents histograms of distance to boundary for land (a) and sales (b). Distance to boundary is in feet, with positive distance indicating being inside the flood zone. Excludes boundaries that trace a body of water and pixels that overlap with the boundary.

Figure A-3: RD Estimates: Current Flood Zone Status

(a) Share of Land in 1996 Flood Zone



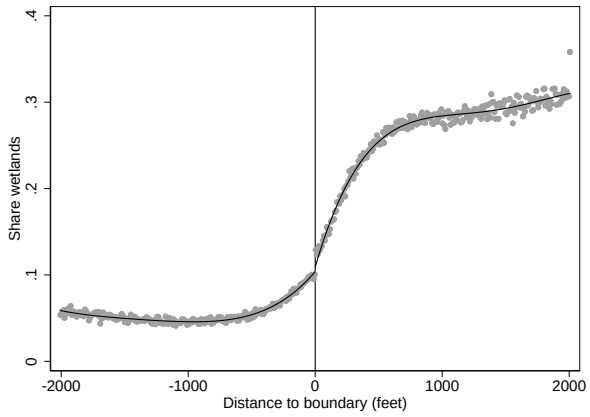
(b) Share of Land in 2017 Flood Zone



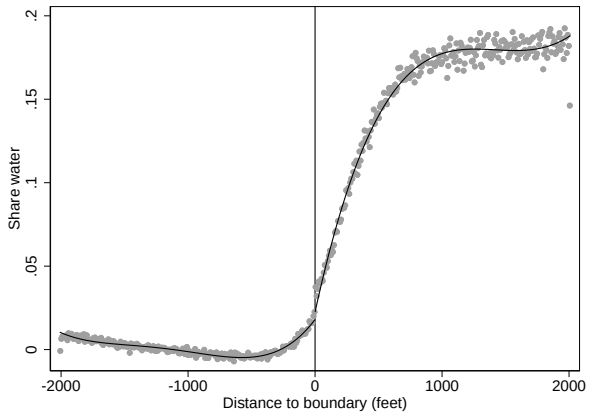
Notes: Figures present RD plots with a fourth order polynomial fit on either side of the flood zone boundary. Distance to boundary is measured in feet, with positive distance indicating being inside the flood zone. Sub-figure (a) plots the share of land in the 1996 flood zone, and sub-figure (b) plots the share of land in the 2017 flood zone. Estimates are residualized of census tract fixed effects.

Figure A-4: RD Estimates: Other Pre-Period Land Use

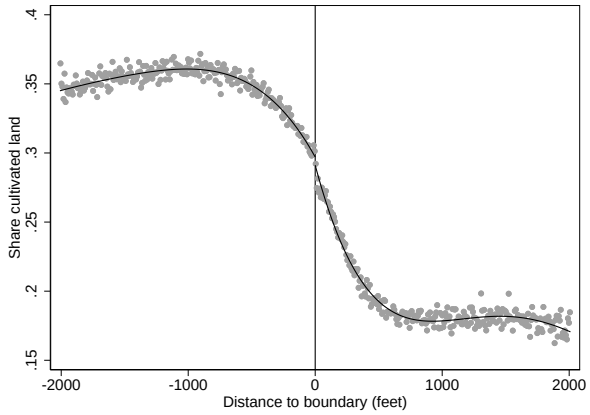
(a) Wetlands



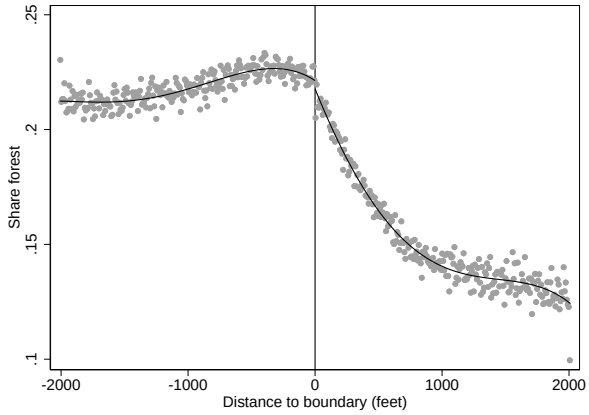
(b) Water



(c) Agriculture



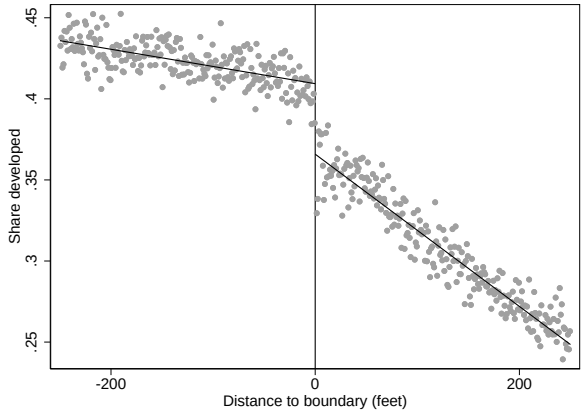
(d) Forest



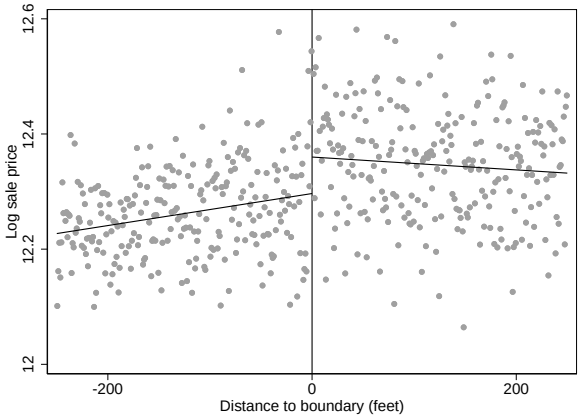
Notes: Figures present RD plots with a fourth order polynomial fit on either side of the flood zone boundary. Distance to boundary is measured in feet, with positive distance indicating being inside the flood zone. All land use outcomes are measured as of the late 1970s and early 1980s. Estimates are residualized of census tract fixed effects.

Figure A-5: RD Figures: Local Linear Specification

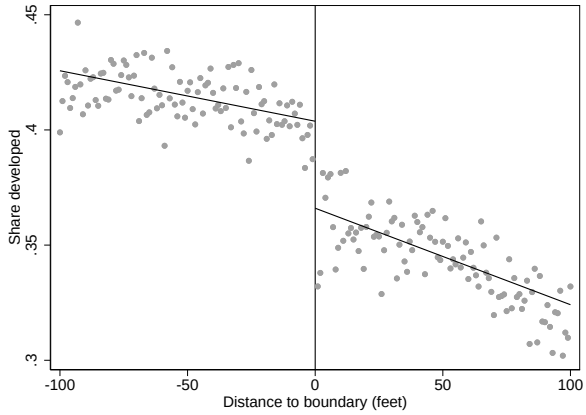
(a) Share Developed in 2016, Rectangular Kernel



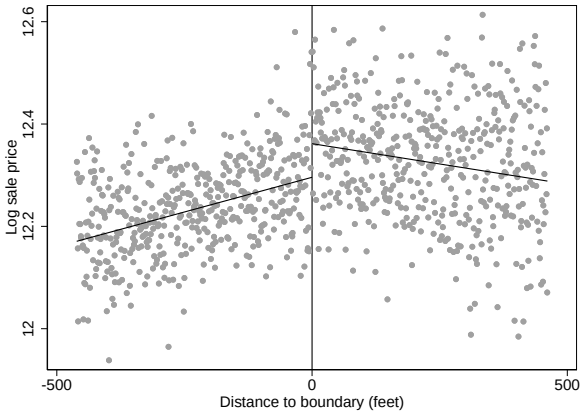
(b) Log Sales Price, Rectangular Kernel



(c) Share Developed in 2016, Triangular Kernel

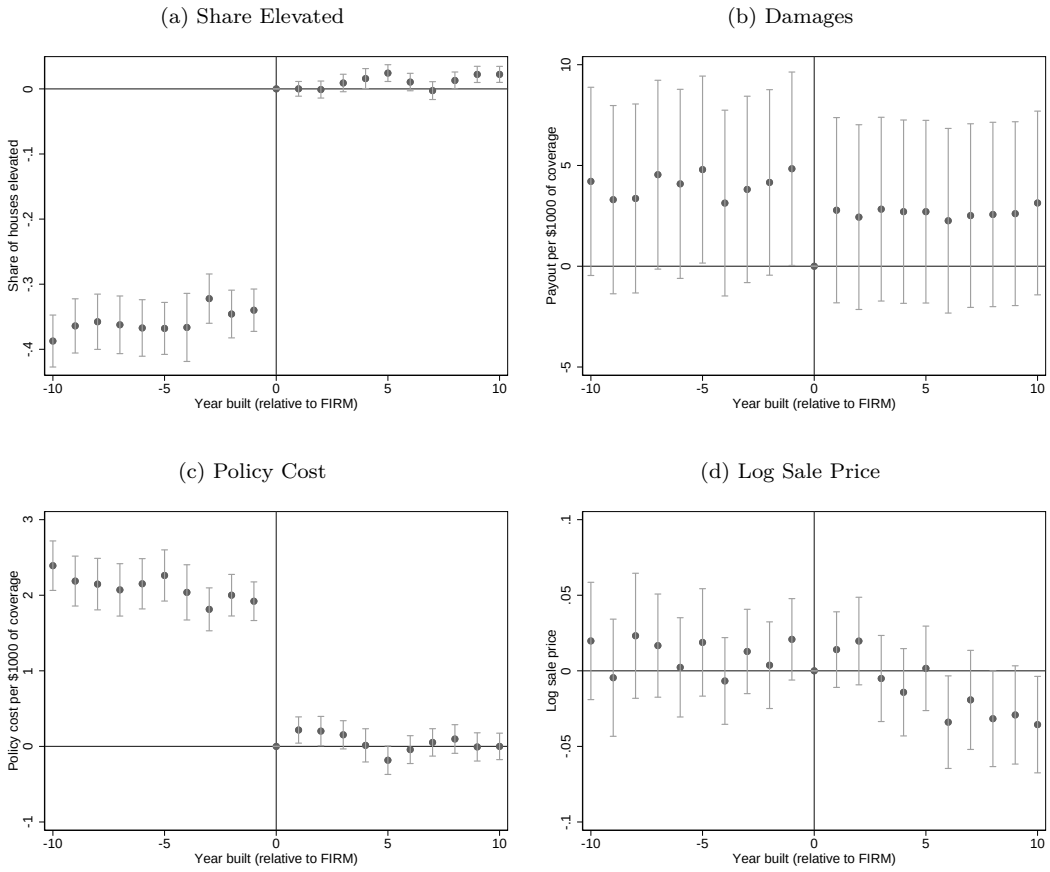


(d) Log Sales Price, Triangular Kernel



Notes: Figures present RD plots for alternative specifications. Sub-figures (a) and (b) present results using a local linear regression with a rectangular kernel on a fixed 250 foot bandwidth. Subfigures (c) and (d) present results using a local linear regression with a rectangular kernel on the MSE-optimal bandwidth following Calonico et al. (2014). Estimates are residualized of census tract fixed effects.

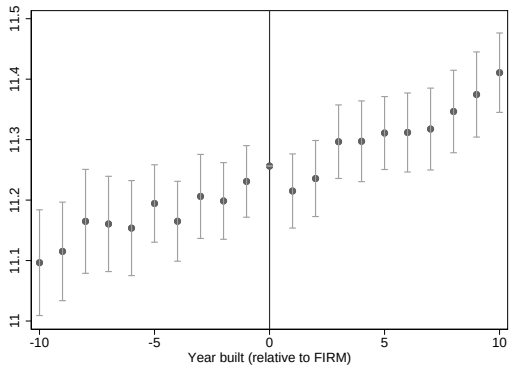
Figure A-6: NFIP Enrollment Year Difference-in-Difference Estimates



Notes: Figures present coefficients from the difference-in-difference specification on bins of year built relative to FIRM year (the year of enrollment in the National Flood Insurance Program, at which time building codes began to be imposed on newly-constructed housing). Sample is restricted to single-family residences.

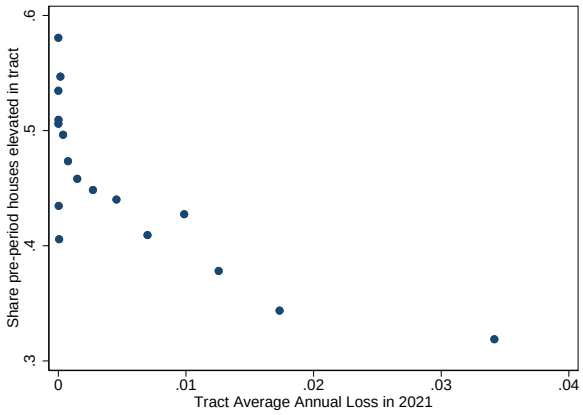
Figure A-7: NFIP Enrollment Year Event Study Estimates of Residualized Sales Price

(a) Log Sale Price



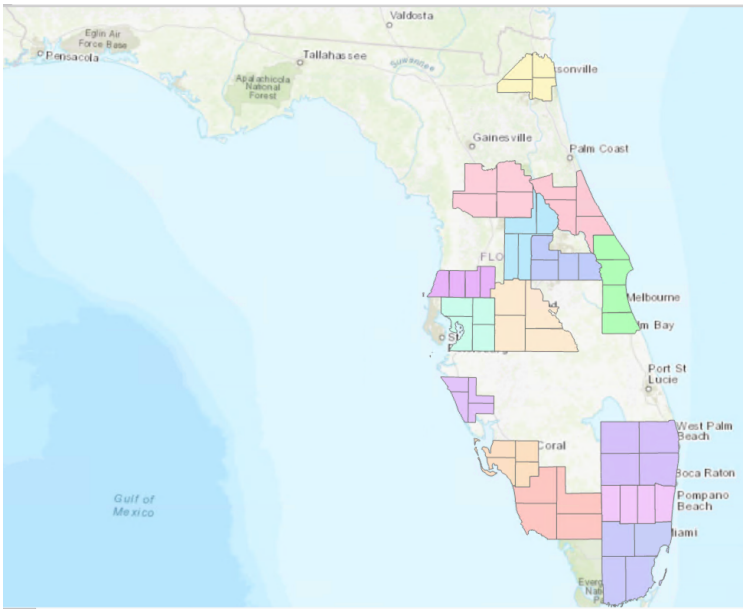
Notes: Figures present coefficients on bins of year built relative to FIRM year (the year of enrollment in the National Flood Insurance Program, at which time building codes began to be imposed on newly-constructed housing). Outcome is sales price, residualized of fourth-degree polynomials in parcel size and total living area, and county-by-sale-month and year built fixed effects. Sample is restricted to properties inside the flood zone.

Figure A-8: Tract Share of Pre-Regulation Houses Elevated, by Flood Risk



Notes: Figure plots a binned scatter plot of the share of houses that were built before building code regulations in each tract, against that tract's average annual loss in 2021. Sample is restricted to houses inside the flood zone.

Figure A-9: Quadrants



Notes: Figure depicts the top 16 counties with most development in Florida, divided into equal-area quadrants.

Table A.1: Tabulating discrepancies between flood zone status and the First Street Model

	Equal weight to each developed pixel		Weighted by number of parcels	
	Inside flood zone	Outside flood zone	Inside flood zone	Outside flood zone
	(1)	(2)	(3)	(4)
Inside First Street 100 year floodplain	0.188	0.136	0.277	0.104
Outside First Street 100 year floodplain	0.093	0.583	0.136	0.484

Notes: Table tabulates the share of all buildings in our eleven counties of interest that fall into each of the four mutually exclusive categories of flood zone status by First Street floodplain status. Flood zone status designates areas that FEMA has determined are in a 100 year floodplain (i.e. they have a greater than 1 percent chance of flooding per year). Columns (1) and (2) tabulate the share of pixels covered by a building footprint that are in each category. Columns (3) and (4) tabulate the number of parcels (accounting for multiple parcels on the same pixel).

Table A.2: Share Mapped Out of Floodplain by Land Use

	Inside flood zone in 2004	
	Land Share of Flood Zone (2004)	Share Mapped Out of Flood Zone in 2008/2009
	(1)	(2)
Any Developed	0.148	0.230
Developed - Open	0.045	0.160
Developed - Low	0.056	0.224
Developed - Mid	0.032	0.282
Developed - High	0.015	0.351
Wetlands	0.621	0.004
Water	0.064	0.003
Cultivated	0.076	0.068
Other Land Use	0.092	0.059

Notes: Table presents the share of the land inside the flood zone in 2004 that is remapped out of the flood zone in the next remapping, by land use category in 2004. Sample is restricted to Marion and Dade Counties, because these were the two counties in our sample of interest that experienced zero remappings between 1996 and 2004 and one remapping between 2004 and 2016. These restrictions were made so that (1) we could be confident that the flood zone status in 2004 was reflected in the digitized flood maps from 1996 (our next-most-recent set of flood maps) and (2) we could observe a change in flood zone status between 2004 and our modern-day flood maps in 2017. These two counties had maps drawn in 1995 and then 2008 or 2009. We measure land use in 2004 from the NLCD. We compute the share of land inside the flood zone in 2004 that was remapped outside of the flood zone by 2017, splitting by land use category in 2004.

Table A.3: RD Analysis Results: Other Land Use Outcomes

	Polynomial		Local Linear (CCT bandwidth)	
	Pre-period	Current	Pre-period	Current
	(1)	(2)	(1)	(2)
Wetlands	0.005	0.110	0.019	0.071
	(0.004)	(0.007)	(0.003)	(0.006)
Water	0.005	-0.020	0.012	0.001
	(0.003)	(0.004)	(0.002)	(0.001)
Agriculture	-0.006	-0.029	-0.015	-0.021
	(0.004)	(0.007)	(0.003)	(0.005)
Forest	-0.004	-0.009	-0.006	0.001
	(0.003)	(0.006)	(0.003)	(0.004)

Notes: Table displays estimates of equation (1) from a fourth order polynomial, fit separately on either side of the boundary, restricted to a window of 2,000 feet on either side of the boundary. All discontinuities are estimated on the historic boundaries and exclude boundaries that trace a body of water. Robust standard errors are clustered at the census tract level.

Table A.4: RD Estimates: Other Sale Price Estimates and Compositional Differences

	Log sale price				Log square footage
	Baseline	Residualized of characteristics	Homes built pre regulations (pre-FIRM)	Homes built post regulations (post-FIRM)	Baseline
	(1)	(2)	(3)	(4)	(5)
All houses	0.065	0.057	0.070	0.048	0.057
	(0.022)	(0.067)	(0.030)	(0.021)	(0.021)
Single-family houses	0.056	0.093	0.103	0.038	0.024
	(0.018)	(0.066)	(0.039)	(0.025)	(0.010)

Notes: Table displays estimates of equation (1) using a fourth order polynomial, fit separately on either side of the boundary, restricted to a window of 2,000 feet on either side of the boundary. Column (2) presents results using the residuals of a regression of log sales price on polynomials in land square footage and interior square footage, sale date (month and year) by county fixed effects, and year built fixed effects. Columns (3) and (4) present results for homes that were built pre vs. post the introduction of the NFIP and the enforcement of building regulations. Column (5) runs the RD regression with the outcome of log square footage of the home conditional on being built. All discontinuities are estimated on the historic boundaries and exclude boundaries that trace a body of water. Robust standard errors are clustered at the census tract level.

Table A.5: Summary Statistics of Construction Year RD Sample

	Inside flood zone		Outside flood zone	
Variable	(1) Pre-Enrollment	(2) Post-Enrollment	(3) Pre-Enrollment	(4) Post-Enrollment
Share elevated	0.562	0.927		
Building coverage	$188,970	$203,384	$200,232	$203,867
Contents coverage	$42,963	$51,360	$78,908	$80,500
Policy cost	$1,058	$658	$438	$433
Payout	$444	$187	$240	$222
N policy-years	598,961	618,265	236,628	391,829
House price (2010 $USD)	$263,253	$307,565	$166,752	$188,521
N house sales	69,976	103,620	147,663	304,083

Notes: Table presents variable means in the estimation sample for the analysis of the effect of building codes on elevation, insurance payouts, premiums, and house prices. Elevation, payout, and cost data come from residential NFIP policies from 2010-2018. Price data come from residential sales prices in 2005-2020. We use all single-family residences in Florida. Sample is restricted to houses constructed +/- 10 years around NFIP enrollment.

Table A.6: Summary Statistics of the Model Estimation Sample

	Whole Sample		Balanced Boundary Sample (est. samp.)	
	(1)	(2)	(3)	(4)
	Outside flood zone	Inside flood zone	Outside flood zone	Inside flood zone
Share developed, 1980	0.65	0.49	0.56	0.50
N developed gridcells, 1980	2,117	884	94	74
Share developed, 2016	0.84	0.68	0.81	0.73
N developed gridcells, 2016	5,820	1,960	240	179
Share elevated pre-regulation	0.00	0.39	0.00	0.45
Share elevated post-regulation	0.00	1.00	0.00	1.00
Home price, 1980	$46,405	$51,510	$49,785	$49,785
Home price, 2017	$207,555	$302,264	$238,150	$282,524
First Street AAL (expected annual damages in 2021, middle scenario)	0.0005	0.0042	0.0007	0.0019
First Street AAL (expected annual damages in 2051, middle scenario)	0.0013	0.0084	0.0019	0.0043
N gridcells	20,254	11,231	564	527
N observations	1,045	805	254	254

Notes: Table presents summary statistics of the aggregated sample at the tract-zone-boundary proximity level. Columns (1) and (2) describe the whole sample, used for counterfactuals. Columns (3) and (4) describe the subset of the sample used for estimating the coefficients of interest. This subset is restricted to paired inside/outside flood zone observations that are within 100 feet of a boundary. Each observation has the same weight regardless of share developed.

Table A.7: Estimated Parameters, Alternative Specifications

	Specification			
	(1)	(2)	(3)	(4)
Supply cost of flood zone designation (ψ)	0.246	0.246	0.246	0.246
	(0.296)	(0.322)	(0.286)	(1.603)
Demand cost of flood zone designation ($-\phi$)	0.230	0.269	0.083	0.247
	(0.491)	(0.572)	(0.159)	(0.813)
Demand elasticity (α^D)	-1.0	-0.4	-3.3	-0.7
				(1.041)
Consumer WTP to avoid flood zone (ϕ/α^D)	0.230	0.672	0.025	0.333
Estimating α^D?	N	N	N	Y

Notes: Table presents parameter estimates from our baseline specification. Column (1) uses a demand elasticity calibrated to -1 based on Song (2021). Columns (2) and (3) use demand elasticities at either end of the range imputed from estimates in Calder-Wang (2021). Column (4) estimates the demand elasticity in-sample using residual variation (i.e. not accounting for potential endogeneity between price and unobserved quality). Standard errors (in parentheses) were generated from bootstrapping (100 iterations).

A.2 Detailed Methods

A.2.1 Data Processing

Algorithm for Selecting Counties to Digitize

The historic flood maps are available online in a series of scanned images. These maps are organized first by county and then by "community," which can be as small as a village or as large as all unincorporated areas of a county. Each community is mapped in a series of tiles. Tiles vary in size and amount of land covered. Because we faced a fixed per-tile digitization cost, and we had a limited budget, our goal was to select the fewest number of tiles that give the most useful variation. In particular, we wanted to ensure we digitized tiles that saw substantial development between the 1980s and present day, but focused on sufficiently large areas to avoid any concerns about selecting on an endogenous outcome.

Our process for selecting maps was as follows:

Step 1. Select the top 15 counties with the largest quantity of newly-developed land, according to our digitized land use data.[1]

[1]We restricted to the top 15 counties because each county requires substantial manpower to evaluate (we had to manually determine the location of each tile in order to assign it to a quadrant).

Step 2. Divide each county into equal-area quadrants. An illustration of the quadrants is shown in Figure A-9.

Step 3. For each quadrant, compute total area of new development.

Step 4. For each quadrant, count the number of tiles that overlap it.

Step 5. For each quadrant, compute the total area of new development per tiles that would need to be mapped.

Step 6. Sort quadrants by area of new development per tile and drop quadrants with the lowest value until budget constraint is met.

We ended up with 120 tiles from 11 different counties (21 quadrants). An alternative procedure, in which we first dropped all quadrants with more than ten tiles, and then selected the quadrants with largest total area of new development, yielded a very similar set of quadrants.

A.2.2 Computing Distance to Boundaries

We compute the distance from each grid cell to the closest point on a flood zone boundary that is not within 100 feet of the border of a body of water. We also compute the distance to county boundaries. We drop grid cells that fall within 2 miles of county boundaries to avoid accidentally including flood zone boundaries that overlap county boundaries.

A.2.3 NFIP Enrollment Year Analysis: Additional Material

Data Restrictions

We restrict to residential policies on single-family homes and drop any policies whose coverage exceeds statutory caps. We measure payouts as the total claims paid out for building, contents, and total cost of compliance insurance.[2] We measure policy cost as the total of the premium and other fees. We measure elevation using an indicator of whether a house's elevation exceeds the base flood elevation (BFE).[3] By definition, this variable is not available outside the floodplain since these areas are assumed to be above the base flood elevation. Inside the floodplain, it is available in about half of pre-period (unregulated) houses and almost 100% of post-period (regulated) houses. We measure a house as elevated if the measured difference between lowest floor elevation and BFE is greater than or equal to 0. Pre-period houses can get cheaper policies if they can show they are elevated to the BFE, providing an incentive to report only if they are elevated. Because of this, we assume that if the elevation is missing, the house is not elevated.

To construct the dataset on house prices, we match building footprints to their respective parcels in the Florida parcel data. The parcel data records the effective year built of each house. We then match the

[2]Total cost of compliance pays for rebuilding that is required when a house is more than 50% damaged and is required to meet a higher building standard when it is rebuilt.

[3]The Base Flood Elevation measures the height of the flood that has a greater than 1% chance of happening every year.

centroid of each building footprint to the NLCD gridcell into which it falls. We assign the building footprint the census tract and flood zone status of that NLCD gridcell. We measure sales price as the average arms-length transaction price for single family residences between 2005 and 2020. The next section describes how we measure the year of NFIP enrollment for each census tract.

Determining NFIP Enrollment Year

We define the year of each community's NFIP enrollment, and therefore the year in which floodplain regulations were imposed, using data on NFIP policies from 2009 to 2019. We use the fact that policies have an indicator for whether the house was built post-FIRM to construct the year of NFIP enrollment at the census tract level. Because enrollment occurred at the community level, and a community is generally larger than a census tract, characterizing year of enrollment at the census tract level is unlikely to introduce substantial inaccuracies. We define the year of enrollment as the first year within a census tract in which over 50% of homes are coded as post-FIRM. We restrict to census tracts with at least 25 distinct years of construction to avoid classifying the enrollment year based on noise.

Difference-in-Difference Specification

We expand on our event-study strategy in Section 2.4.2 by also using the fact that building standards were imposed for houses built inside the flood zone but not for houses built outside of it. We estimate the following difference-in-difference models for houses in the flood zone:

$$y_i = \Sigma_r \beta_r 1\{r_i = r\} SFHA_i + \Sigma_r \gamma_r 1\{r_i = r\} + \delta SFHA_i + \gamma_{j(i)} + \varepsilon_i \tag{A.1}$$

$$y_i = \alpha + \beta Post_i SFHA_i + \gamma Post_i + \delta SFHA_i + \nu r_i + \eta r_i Post + \phi r_i SFHA_i + \psi r_i Post_i SFHA_i + \gamma_{j(i)} + \varepsilon_i \tag{A.2}$$

Here, $\gamma_{j(i)}$ indicates census tract-by-flood-zone fixed effects. We again cluster standard errors at the census tract level. In all claims and policy regressions, we weight each observation by the number of policies it represents.

The patterns observed in Section 2.4.2 are largely unchanged in the difference-in-difference specification. Here, we find that the regulation reduces insurance payouts by \$2.84 per \$1000 of coverage. At the average coverage amount of \$252,000, this translates to a difference of \$715 per year. With a discount rate of 5%, the PDV of these savings is 7.5% of the average house value inside and outside the flood zone. Similarly, the difference-in-difference specification finds that the regulation reduces policy premiums by about \$1.78 per \$1000 of coverage, or \$450 per year. The PDV of these savings is about 4.5% of the average house value.

Stylized Model of WTP for Adaptation

Suppose that houses are either adapted (A) or non-adapted (B), with a fixed supply of each. Denote c as the (total lifetime) savings from living in an adapted house ($c < 0$ means less money is spent on premiums) and ρ as the share of savings that are internalized by the home-buyer. Let p_A and p_B be the respective house prices (in levels), α be the demand elasticity for house price, and assume supply of houses is fixed.

Suppose $u_{iA} = \alpha(p_A + \rho c) + \varepsilon_{iA}$ and $u_{iB} = \alpha p_B$ where ε_{iA} is distributed i.i.d Type 1 Extreme Value. This specification embeds the assumption that consumers only care about adaptation through its effects on risk. The distribution of ε_{iA} yields that the share of adapted houses $s_A = \frac{\exp(\alpha(p_A - p_B + \rho c))}{1+\exp(\alpha(p_A - p_B + \rho c))}$. Algebraic manipulation reveals that given the prices, the elasticity of demand, and the shares, we can compute the share of internalized savings as

$$\rho = \frac{1}{c}\left(\frac{1}{\alpha}\ln\left(\frac{s_A}{1-s_A}\right) - (p_A - p_B)\right)$$

We assume that $\alpha = -1$ and take $s_A = 0.8$ based on the market share of post-FIRM houses in 2016, and $c = -6398$ based on our estimates. An estimate of no price difference between adapted and non-adapted houses yields an internalized share $\rho$of approximately zero.

The upper end of the 95% confidence interval is a price increase of 1.06%. This would translate to 35% of the PDV of insurance payout savings and 46% of the PDV of insurance premium savings. If we instead apply this upper-bound estimate we find that $\rho = 0.36$. That is, consumers only internalize a maximum of 36% of the PDV of the savings that would accrue to them from an adapted house.

A.2.4 Calibration of Supply Elasticities

We use as our starting point estimates produced in Baum-Snow and Han (2022) (BSH) for land development elasticities at the census tract level, estimated between 2001 and 2011. Specifically, we take estimates of the elasticities as of 2001 from the IV specification. We replace any negative elasticities with the smallest nonnegative elasticity in our eleven-county sample. We denote these elasticities as α_j^S.

For each tract, we average the 2016 price and quantity within all observations in the tract, weighting by number of grid cells. We then compute the decrease in share developed that would be implied by the BSH elasticities for a decrease of 10% in the price from the observed 2016 price: $\tilde{q}_j = q_j^{2016} - \alpha_j^S(0.1p_j^{2016})$.[4] We then use our postulated relationship that

$$q_j^{2016} = \frac{Q_j^{2016}}{L_j} = \Phi\left(\frac{p_j^{2016} - \psi E_j^{2016} - \mu_j - \eta_j}{\sigma_j}\right)$$

to recover μ_j and σ_j under the assumption that $\eta_j = 0$ and E_j^{2016} is the elevation status in the tract in 2016. We compute σ_j and μ_j as

[4]We use 2016 price and quantity as our baseline because the BSH estimates used data from 2000-2010.

$$\sigma_j = \frac{-0.1 p_j^{2016}}{\Phi^{-1}(\tilde{q}_j) - \Phi^{-1}(q_j^{2016})} \tag{A.3}$$

$$\mu_j = p_j^{2016} - \psi E_j^{2016} - \Phi^{-1}(q_j^{2016})\sigma_j \tag{A.4}$$

Since ψ enters the calculation of μ_j and σ_j , we follow a two-step procedure where we first calculate μ_j and σ_j with the assumption that $\psi = 0$, then we re-estimate using the estimated value of ψ.

A.2.5 Moment Condition Details

Demand-side identification and moments

We assume that in a narrow band around the floodplain boundary, some portion of unobserved amenities $\Delta^D SFHA_z$ is a function of flood zone status, but the remainder $\xi_{jz} - \Delta^D SFHA_z = \tilde{\xi_{jz}}$ is uncorrelated with flood zone. For a given value of Δ^D, we compute the "de-biased" price $\tilde{p}_{jz} = \delta_{jz} - \alpha^D p_{jz}^{2016} - \phi SFHA_z - \tilde{\xi}_{jz}$. We then compute the difference in "de-biased" prices at the boundary as:

$$\begin{aligned}\tilde{p}_{j1} - \tilde{p}_{j0} &= \frac{1}{\alpha^D}\left(\delta_{j1} - \delta_{j0} - \phi - (\tilde{\xi}_{j1} - \tilde{\xi}_{j0})\right) \\ &= \frac{1}{\alpha^D}\left(\delta_{j1} - \delta_{j0} - \phi - (\xi_{j1} - \Delta^D - \xi_{j0})\right)\end{aligned}$$

and we set the expected difference in price across the boundary equal to $\beta^{p,2016}$. This yields the moment:

$$E\left[\frac{1}{\alpha^D}\left(\delta_{j1} - \delta_{j0} - \phi - (\xi_{j1} - \Delta^D - \xi_{j0})\right) - \beta^{p,2016}\right] = 0 \tag{A.5}$$

The non-correlation between $\tilde{\xi_{jz}}$ and flood zone yields the following moments, where we introduce μ_D as a constant baseline unobserved amenity level:

$$E[(\tilde{\xi_{jz}} - \mu_D)SFHA_z] = 0 \tag{A.6}$$

$$E[(\tilde{\xi_{jz}} - \mu_D)] = 0 \tag{A.7}$$

Supply-side calibration and moments

We assume that in a narrow band around the floodplain boundary, some portion of unobserved construction costs $\eta_{jz}^t = \Delta^{S,t} SFHA_z$ are a function of floodplain status, while the remainder $\tilde{\eta}_{jz}^t = \eta_{jz}^t - \Delta^{S,t} SFHA_z$ are

uncorrelated with floodplain status. In each period, for a given value of $\Delta^{S,t}$, we compute the "de-biased" share developed $\Phi\left(\frac{p_{j1}^t - \psi E_{j1}^t - \mu_j - \tilde{\eta}_{j1}^t}{\sigma_j}\right)$. We then require the difference in "de-biased" share developed at the boundary to match the RD quantity coefficients $\beta^{q,1980}$ and $\beta^{q,2016}$. The moments that identify $\Delta^{S,1980}$ and $\Delta^{S,2016}$ are:

$$E\left[\Phi\left(\frac{p_{j1}^{1980} - \psi E_{j1}^{1980} - \mu_j - \tilde{\eta}_{j1}^{1980}}{\sigma_j}\right) - \Phi\left(\frac{p_{j0}^{1980} - \psi E_{j0}^{1980} - \mu_j - \tilde{\eta}_{j0}^{1980}}{\sigma_j}\right) - \beta^{q,1980}\right] = 0 \tag{A.8}$$

$$E\left[\Phi\left(\frac{p_{j1}^{2016} - \psi E_{j1}^{2016} - \mu_j - \tilde{\eta}_{j1}^{2016}}{\sigma_j}\right) - \Phi\left(\frac{p_{j0}^{2016} - \psi E_{j0}^{2016} - \mu_j - \tilde{\eta}_{j0}^{2016}}{\sigma_j}\right) - \beta^{q,2016}\right] = 0 \tag{A.9}$$

We then recover the supply cost of regulation ψ and the baseline supply cost terms $\mu_{s,1980}$ and $\mu_{s,2016}$ with the moments:

$$E[(\tilde{\eta}_{jzb}^{1980} - \mu_{s,1980}) SFHA_z 1\{b = close\}] = 0 \tag{A.10}$$

$$E[(\tilde{\eta}_{jzb}^{2016} - \mu_{s,2016}) SFHA_z 1\{b = close\}] = 0 \tag{A.11}$$

$$E[(\tilde{\eta}_{jzb}^{1980} - \mu_{s,1980}) 1\{b = close\}] = 0 \tag{A.12}$$

$$E[(\tilde{\eta}_{jzb}^{2016} - \mu_{s,2016}) 1\{b = close\}] = 0 \tag{A.13}$$

A.2.6 Details on Data for Model Estimation

We measure today's sales price p_{jz}^{2016} with the log of the median sales price for single-family homes from 2014-2019 based on the location of the building footprint. We measure sale price in the pre-period p_{jz}^{1980} as the log of the median value of owner-occupied non-condominium housing units from the 1980 Census. These data are not available at the flood zone level. Because floodplains did not exist in the pre-period we assume that the price does not differ between floodplains within a Census tract. We measure quantity of developed land in 1980 and 2016 as the number of gridcells that are categorized as developed in the 1980 and 2016 land-use datasets.[5] We measure elevation from NFIP policy data, as discussed in Appendix A.2.3. We define a tract as elevated if more than 50% of insured houses in that tract are elevated. Where we do not observe elevation (including all non-flood-zone tracts), we assume it only occurred when required by regulation. When historic flood zone status is unavailable because of the limited reach of our digitized maps,

[5]To account for the fact that some locations jz have no developed land in 1980, we calculate share developed $q_{jz} = (Q_{jz} + 1)/L_{jz}$ in those locations. Also, in the spirit of Burchfield et al. (2006) we correct for potentially mismeasured growth by measuring the number of developed cells in 1980 as the minimum of the observed number of developed cells in the location in 1980 and the number measured in 2016.

we use the 1996 flood zone status to calculate mean utility δ_{jz}, but we restrict to historic boundaries in our estimation of the main parameters.

Appendix Table A.6 presents summary statistics for the model estimation sample. A larger share of our estimation sample is developed than the sample used in Section 2.4,but house prices and flood risk look similar in the two samples.

A.2.7 Calculation of Expected Damages

We define flood risk using data from the First Street Flood Lab estimates of Average Annual Loss (AAL). AAL expresses expected annual damages as a share of house price. These data come from parcel-specific estimates (as opposed to the raw hazard layer) that combine the raw hazard layer (which generates the parcel-specific inundation depth) with the output of an engineering damage model. The damage model takes as inputs a number of features of the structure, including its market value, number of stories and units, and foundation type, and calculates damages using the HAZUS-MH methodology. The HAZUS-MH methodology was developed for FEMA to calculate estimated damages from natural disasters and is based on a set of depth-damage curves collected from FEMA's Federal Insurance and Mitigation Administration (FIMA) and the USACE Institute for Water Resources (USACE-IWR).[6] Average annual loss is expected to grow over time; we assume risk increases linearly from the estimated 2021 risk to the estimated 2051 risk and then stays constant at the 2051 risk for all future years. Wherever First Street did not provide an AAL estimate but did provide a Flood Factor (another measure of risk), we assumed the AAL was 0.

Expected damages are computed as the product of number of newly-developed gridcells and the PDV of expected damage under a given counterfactual. The expected damage is computed as $0.7 \times P_{jz}^{Obs} \times AAL_{jz} \times M_{jz}^{CF}(E_{jz}^{2016})$, where P_{jz}^{Obs} is the observed (level) price of a house and AAL_{jz} is the observed average annual loss.[7] The term M_{jz}^{CF} is a multiplier that accounts for differences in elevation in each counterfactual. We assume that if an observation was elevated in the pre-period it will be elevated in the post-period for all counterfactuals. Otherwise, houses that are not observed to be elevated in the post-period but are elevated in a counterfactual will experience 55% lower damages in that counterfactual; a similar calculation applies to houses that are observed to be elevated but are counterfactually non-elevated.

We compute three measures of damages:

$D^{All} = \frac{1.05}{.05} N^{CF} D^{CF}$

$D^{Reloc} = \frac{1.05}{.05} N^{CF} D^{Obs}$

$D^{Adapt} = \frac{1.05}{.05} N^{Obs} D^{CF}$

where N^{CF} denotes the number of newly-developed gridcells under counterfactual CF and D^{CF} denotes the PDV of expected damage under counterfactual CF. The first measure ("all damages") measures total expected damages by multiplying the counterfactual number of newly-developed houses in each area by the expected damages in that counterfactual. The second measure ("relocation-based damages") holds the PDV

[6] See https://assets.firststreet.org/uploads/2021/02/The_Cost_of_Climate_FSF20210219-1.pdf for more details.

[7] This 70% factor was recommended by First Street, who provided the underlying AAL data.

of expected damage constant at the observed level and only changes the number of houses in each location. This captures damages attributable to the location of houses only. The third measure ("elevation-based damages") holds the number of houses in each location constant at the observed level and only changes the expected damage in each location. We then compute per-house damages by dividing the total expected damages of each type by the number of newly-developed houses (which is constant across counterfactuals).

A.2.8 Welfare Calculation Details

We compute consumer surplus differences in each counterfactual scenario relative to the unregulated benchmark. Following Small and Rosen (1981), we calculate per-person consumer surplus in each market m as:

$$CS_i = \frac{-1}{\alpha^D} \ln \Sigma_{j \in J_m, z \in \{0,1\}} \exp(\alpha^D p_{jz} + \phi SFHA_{jz} + \xi_{jz}) \tag{A.14}$$

where j denotes census tract and z indicates flood zone status. For each market, we compute the change in level price required to make per-person consumer surplus in the counterfactual equivalent to that of the same market in the unregulated benchmark. That is, we solve for ΔP_m^{CF} such that

$$\Sigma_{j \in J_m, z \in \{0,1\}} \exp(\alpha^D p_{jz}^{NoSFHA} + \xi_{jz}) = \Sigma_{j,z} \exp(\alpha^D \ln(P_{jz}^{CF} + \Delta P_m^{CF}) + \phi SFHA_{jz} + \xi_{jz}) \tag{A.15}$$

where P_{jz}^{CF} is the house price in levels in the counterfactual of interest, and p_{jz}^{NoSFHA} is the house price in logs in the unregulated benchmark. We then compute differential consumer surplus of new development as the sum of price-equivalents multiplied by the number of new houses in each county N_m: $\Delta CS^{CF} = \Sigma_m N_m \Delta P_m^{CF}$.

We compute government revenue from the tax policy by adding up all taxes levied on newly-developed houses. We compute government revenue under the flood zone policy by estimating the total amount of insurance premiums. Using our flood insurance policy data, we assume that policies inside the flood zone cost \$1484 per year and policies outside the flood zone cost \$572 per year. Applying back-of-envelope calculations to recent estimates of take-up in Florida (Lingle and Kousky, 2018) and inside and outside the floodplain nationally (Bradt et al., 2021), we assume take-up is 45% inside the flood zone, 6% outside the flood zone in high-risk areas (areas with positive probability of flooding more than 2 feet in the 100-year flood), and 0% outside the flood zone in low-risk areas. As with the tax revenue, we calculate premium revenue only for new development.

Appendix B

Appendix for Chapter 3

B.1 Tables and Figures

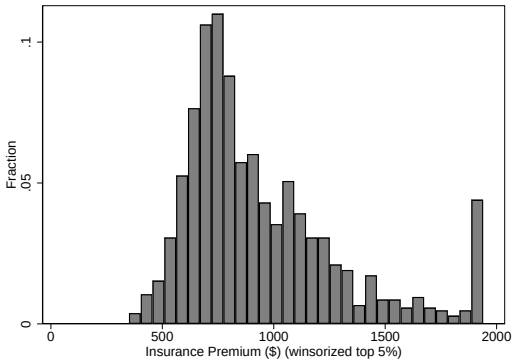
Table B.1: ZIP Code Summary Statistics, All Years

	All ZIPs		Top Quartile Avg Loss Ratio in 2017		Bottom 3 Quartile Avg Loss Ratio in 2017	
	Mean	SE	Mean	SE	Mean	SE
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Insurance Data						
ZIP Loss Ratio in 2017	1.9	0.3	2.3	0.2	1.8	0.2
Premium Price	$1,034	$570	$1,522	$922	$911	$345
N Policies per year	5,481	4,100	4,030	3,322	5,844	4,195
N Policies per owner-occupied unit	1.70	18.44	1.31	2.17	1.79	20.59
Panel B. Housing Market Outcomes						
N Single-family house sales	228.6	214.8	172.2	168.3	242.6	222.7
N Sales per owner-occupied unit	0.062	0.301	0.092	0.584	0.054	0.168
N Mortgage originations	240.0	223.0	168.4	160.8	257.8	232.6
N Mortgage originations per sale	1.17	0.75	1.09	0.82	1.19	0.72
Median single-family house sale price	$741,512	$644,263	$1,207,021	$976,240	$625,134	$461,534
Median sale price per square foot	$409	$264	$565	$358	$370	$218
Panel C. Fire Risk						
Fire Factor Rating	2.6	1.9	2.5	1.4	2.7	2.0
Annual Parcel-level Burn Probability	0.0007	0.0015	0.0004	0.0008	0.0008	0.0016
Wildfire in 2015	0.042	0.201	0.043	0.203	0.042	0.200
Wildfire in 2016	0.055	0.227	0.067	0.250	0.051	0.221
Wildfire in 2017	0	0	0	0	0	0
Wildfire in 2018	0.075	0.263	0.072	0.258	0.075	0.264
Wildfire in 2019	0.065	0.247	0.086	0.281	0.060	0.237
Wildfire in 2020	0.107	0.309	0.134	0.341	0.101	0.301
Panel D. Demographics						
Population in 2010	27,933	21,133	16,748	14,194	30,730	21,653
N owner-occupied housing units	5,079	3,780	3,963	3,207	5,358	3,861
N 1-4 family housing units	7,711	5,386	5,750	4,603	8,201	5,456
Median annual household income	$78,993	$21,610	$91,079	$26,916	$75,972	$18,897
N ZIP Codes	1,045		209		836	

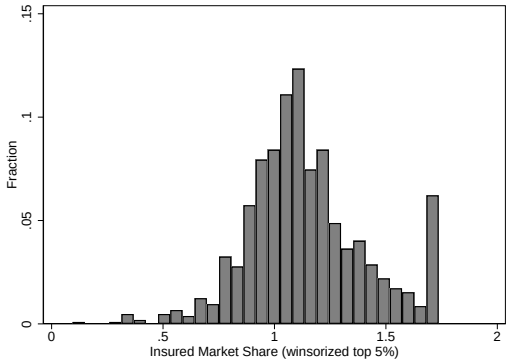
Notes: Table presents summary statistics, for (1) the entire sample, (2) the treatment group, consisting of ZIP codes in the top 25% of average insurer loss ratio in 2017, and (3) the control group, consisting of ZIP codes in the bottom 75% of the distribution. Observations are at the ZIP-year level, for all years.

Figure B-1: Histograms of Primary Outcomes

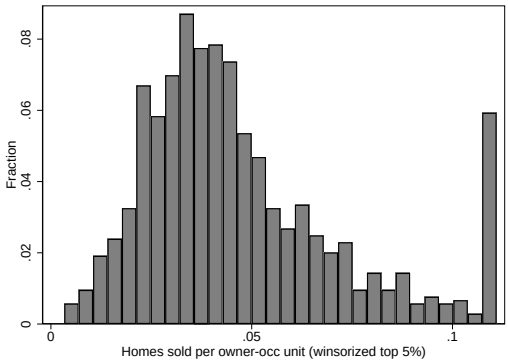
(a) Premiums



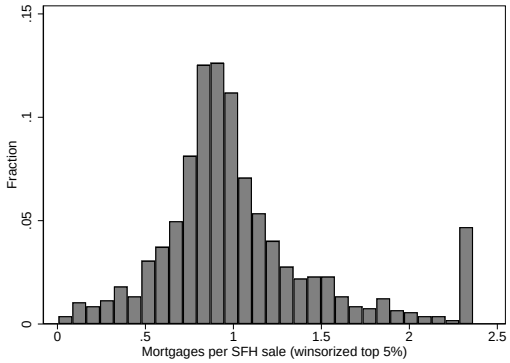
(b) Insured Market Share



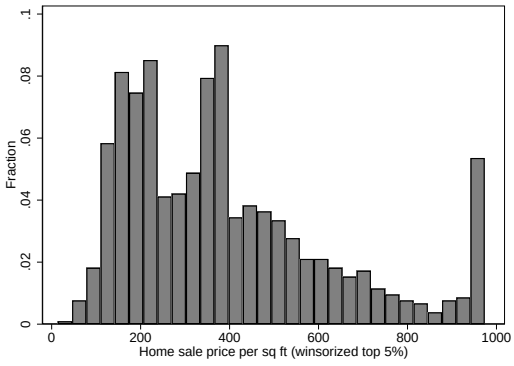
(c) Share Homes Sold



(d) Mortgages per Sale



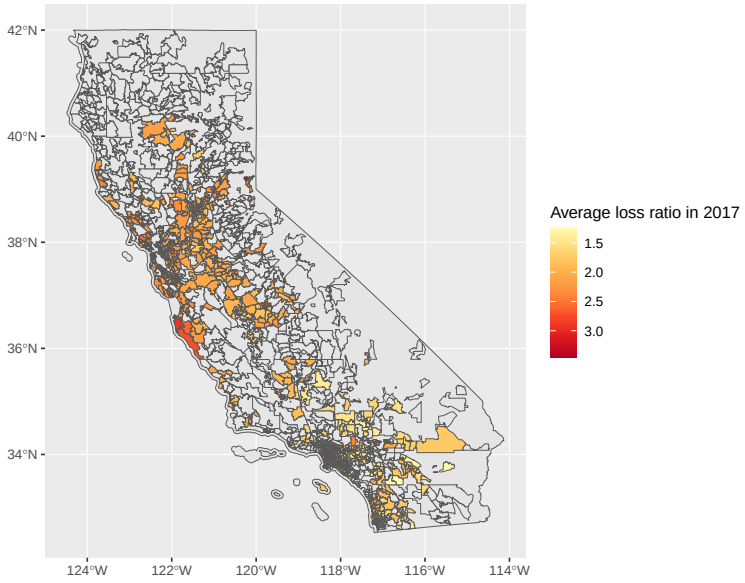
(e) Price per Square Foot



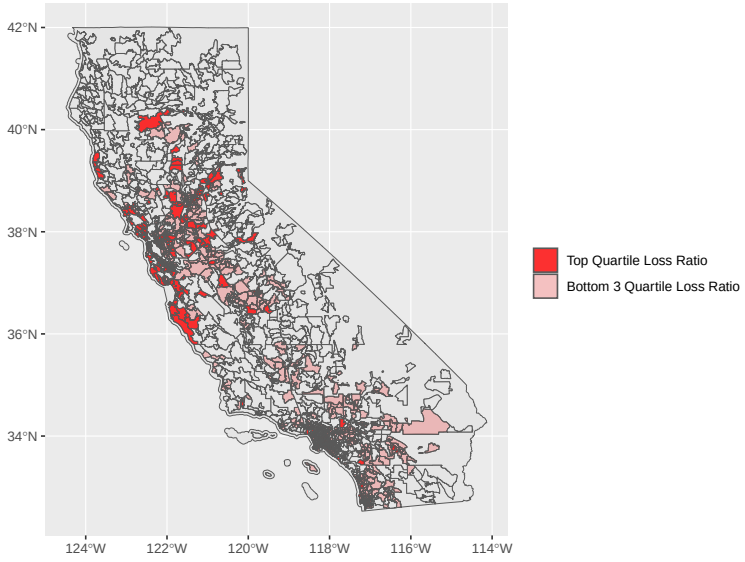
Notes: Figure presents histograms of our main outcomes of interest: earned premiums, insurance policies per owner-occupied house, home sales per owner-occupied house, home sales price per square foot, and number of mortgages per home sale. Each observation is a ZIP code; variables are measured as of 2017.

Figure B-2: ZIP Code Loss Ratio Map

(a) ZIP Code Average Loss Ratios in 2017

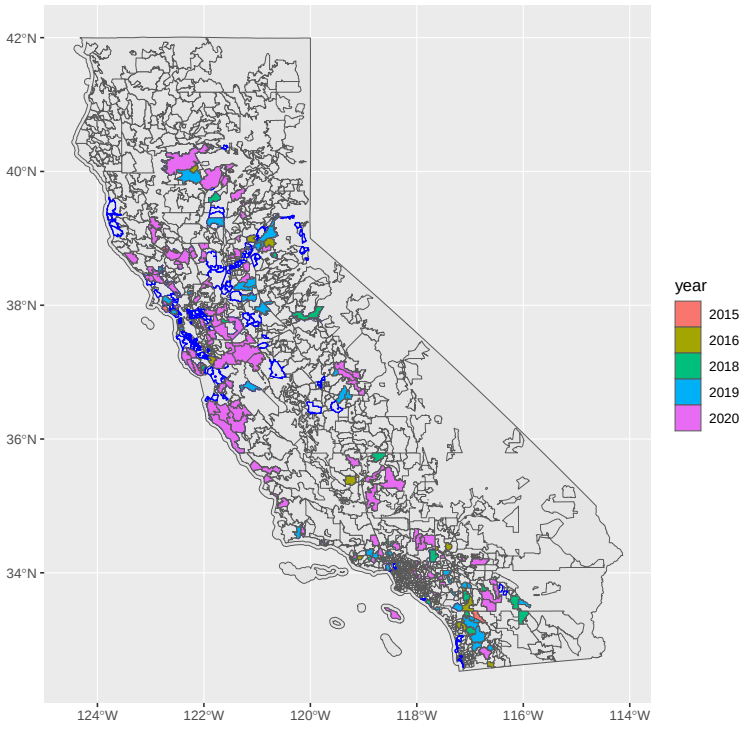


(b) ZIP Codes with High vs Low Loss Ratios



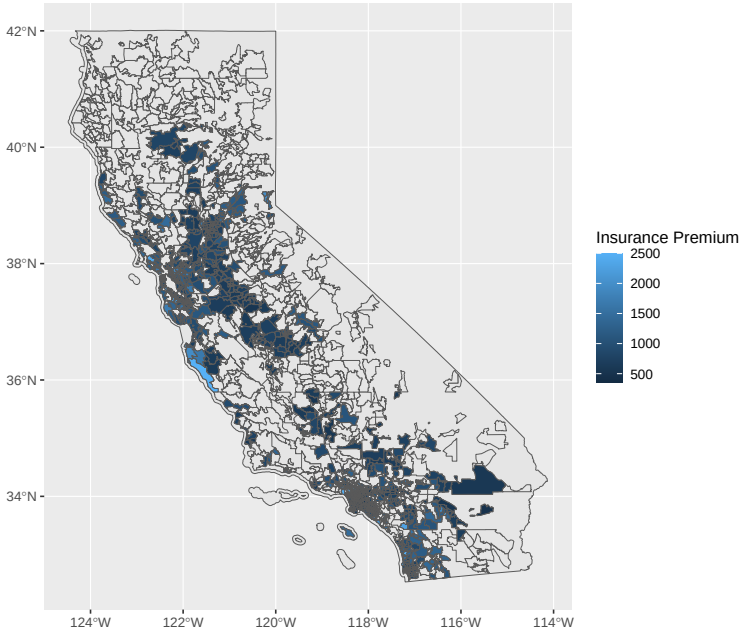
Notes: Figure presents maps of California ZIP codes included in our analysis. Panel (a) illustrates the spatial distribution of loss ratios in 2017, with darker red color indicating a higher loss ratio. Panel (b) presents the same information, with darkest red identifying ZIP codes in the top 25% of loss ratios and lighter red indicating less-intensely-affected ZIP codes.

Figure B-3: ZIP Code Yearly Wildfire Map



Notes: Figure presents a map of California ZIP codes, with colors indicating the year of the latest wildfire. ZIP codes outlined in blue are in the top quartile of average insurer losses in 2017.

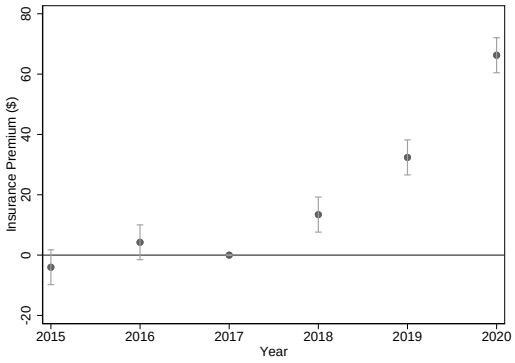
Figure B-4: ZIP Code Insurance Premiums in 2017 Map



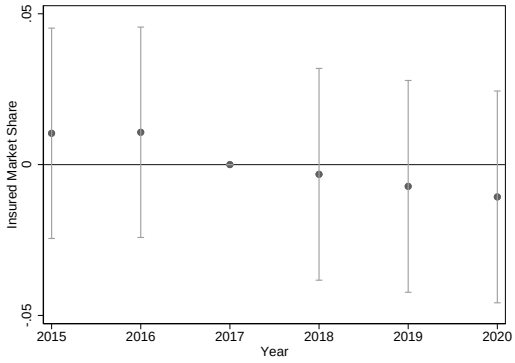
Notes: Figure presents a map of California ZIP codes, with colors indicating the (winsorized) average insurance premium in 2017.

Figure B-5: Pooled Analysis Annual Coefficients

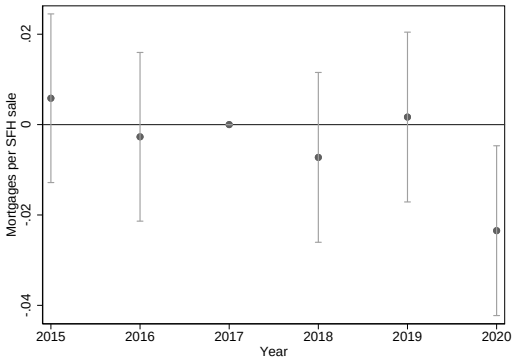
(a) Premiums



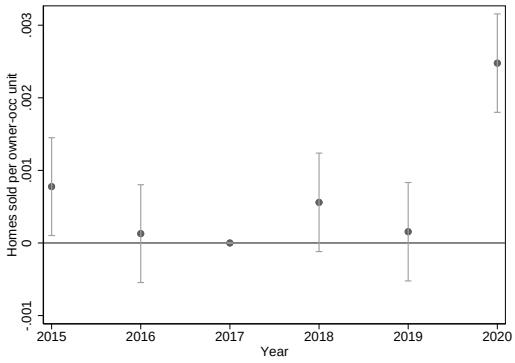
(b) Market Share with Insurance



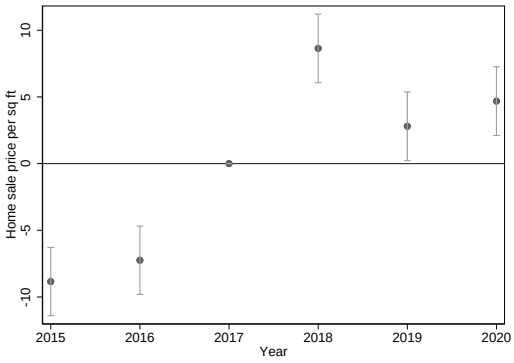
(c) Share of Sales with Mortgage



(d) Share of Homes Sold



(e) Home Sale Price



Notes: Figure depicts estimates of β_t from Equation 3.4 for the outcomes of: (a) average insurance premium, (b) number of insurance policies per owner-occupied housing unit, (c) share of sales with a mortgage origination, (d) share of single-family homes sold, and(e) home sale price per square feet of built property by year.

Bibliography

Allcott, H., and D. Taubinsky, 2015: Evaluating Behaviorally Motivated Policy: Experimental Evidence from the Lightbulb Market. *American Economic Review*, **105 (8)**, 2501–2538, doi:10.1257/aer.20131564, URL https://pubs.aeaweb.org/doi/10.1257/aer.20131564.

American Property Casualty Insurance Association, 2020: Insurance Industry Response to CDI Investigatory Hearing on Homeowners' Insurance Availability and Affordability. Tech. rep. URL https://www.namic.org/pdf/20memberadvisory/1_201020_joint_industry_reply.pdf.

Anagol, S., F. V. Ferreira, and J. M. Rexer, 2021: Estimating the Economic Value of Zoning Reform, URL http://www.nber.org/papers/w29440.

Annan, F., and W. Schlenker, 2015: Federal Crop Insurance and the Disincentive to Adapt to Extreme Heat. *American Economic Review*, **105 (5)**, 262–266, doi:10.1257/aer.p20151031, URL https://pubs.aeaweb.org/doi/10.1257/aer.p20151031.

Bakkensen, L. A., and L. Barrage, 2021: Going Underwater? Flood Risk Belief Heterogeneity and Coastal Home Price Dynamics. *The Review of Financial Studies*, hhab122, doi:10.1093/rfs/hhab122, URL https://academic.oup.com/rfs/advance-article/doi/10.1093/rfs/hhab122/6424922.

Balboni, C., 2019: In Harm's Way? Infrastructure Investments and the Persistence of Coastal Cities, URL http://etheses.lse.ac.uk/3910/1/Balboni___In-harm%27s-way.pdf.

Barahona, N., C. Otero, and S. Otero, 2020: Equilibrium Effects of Food Labeling Policies. *SSRN Electronic Journal*, doi:10.2139/ssrn.3698473, URL https://www.ssrn.com/abstract=3698473.

Baum-Snow, N., and L. Han, 2022: The Microgeography of Housing Supply, URL https://www.atlantafed.org/-/media/documents/news/conferences/2019/12/12/5th-biennial-real-estate-conference/han_the-micro-geography-of-housing-supply.pdf.

Bayer, P., F. Ferreira, and R. McMillan, 2007: A Unified Framework for Measuring Preferences for Schools and Neighborhoods. *Journal of Political Economy*, **115 (4)**, 588–638, doi:10.1086/522381, URL https://www.journals.uchicago.edu/doi/10.1086/522381.

Baylis, P., and J. Boomhower, 2019: Moral Hazard, Wildfires, and the Economic Incidence of Natural Disasters. Working Paper 26550, National Bureau of Economic Research. doi:10.3386/w26550, URL http://www.nber.org/papers/w26550.

Ben-Shahar, O., and K. D. Logue, 2016: The Perverse Effects of Subsidized Weather Insurance. *Stanford Law Review*, **68**, 571–628, URL https://heinonline.org/HOL/P?h=hein.journals/stflr68&i=601.

Bernstein, A., M. T. Gustafson, and R. Lewis, 2019: Disaster on the horizon: The price effect of sea level rise. *Journal of Financial Economics*, doi:10.1016/j.jfineco.2019.03.013, URL http://www.sciencedirect.com/science/article/pii/S0304405X19300807.

Berry, S., J. Levinsohn, and A. Pakes, 1995: Automobile Prices in Market Equilibrium. *Econometrica*, **63 (4)**, 841, doi:10.2307/2171802, URL https://www.jstor.org/stable/2171802?origin=crossref.

Bin, O., and C. 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, URL https://ideas.repec.org/a/eee/jeeman/v65y2013i3p361-376.html.

Blickle, K., and J. a. A. C. Santos, 2022: Unintended Consequences of "Mandatory" Flood Insurance. *SSRN Electronic Journal*, doi:10.2139/ssrn.4086765, URL https://www.ssrn.com/abstract=4086765.

Bradt, J. T., C. Kousky, and O. E. Wing, 2021: Voluntary purchases and adverse selection in the market for flood insurance. *Journal of Environmental Economics and Management*, **110**, 102 515, doi:10.1016/j.jeem.2021.102515, URL https://linkinghub.elsevier.com/retrieve/pii/S0095069621000826.

Brannon, I., and A. Blask, 2017: The government's hidden housing subsidy for the rich. *Politico*, URL https://www.politico.com/agenda/story/2017/08/08/hidden-subsidy-rich-flood-insurance-000495/.

Burby, R. J., 2001: Flood insurance and floodplain management: the US experience. *Environmental Hazards*, **3 (3)**, 111–122, doi:10.3763/ehaz.2001.0310, URL http://www.tandfonline.com/doi/abs/10.3763/ehaz.2001.0310.

Burchfield, M., H. G. Overman, D. Puga, and M. A. Turner, 2006: Causes of Sprawl: A Portrait from Space. *The Quarterly Journal of Economics*, **121 (2)**, 587–633, doi:10.1162/qjec.2006.121.2.587, URL https://academic.oup.com/qje/article-lookup/doi/10.1162/qjec.2006.121.2.587.

Calder-Wang, S., 2021: The Distributional Impact of the Sharing Economy: Evidence from New York City.

Calonico, S., M. D. Cattaneo, and R. Titiunik, 2014: Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs: Robust Nonparametric Confidence Intervals. *Econometrica*, **82 (6)**, 2295–2326, doi:10.3982/ECTA11757, URL http://doi.wiley.com/10.3982/ECTA11757.

Coate, S., 1995: Altruism, the Samaritan's Dilemma, and Government Transfer Policy. *American Economic Review*, **85 (1)**, 46–57, URL https://www.jstor.org/stable/2117995.

CoreLogic, 2019: 2019 Wildfire Risk Report. Tech. rep. URL https://www.corelogic.com/wp-content/uploads/sites/4/downloadable-docs/wildfire-report_0919-01-screen.pdf.

de Moel, H., J. van Alphen, and J. C. J. H. Aerts, 2009: Flood maps in Europe - methods, availability and use. *Natural Hazards and Earth System Sciences*, **9 (2)**, 289–301, doi:10.5194/nhess-9-289-2009, URL https://nhess.copernicus.org/articles/9/289/2009/.

Dell, M., 2010: The Persistent Effects of Peru's Mining Mita. *Econometrica*, **78 (6)**, 1863–1903, doi:10.3982/ECTA8121, URL http://doi.wiley.com/10.3982/ECTA8121.

Deryugina, T., and B. Kirwan, 2018: Does the Samaritan's Dilemma Matter? Evidence from U.s. Agriculture. *Economic Inquiry*, **56 (2)**, 983–1006, doi:10.1111/ecin.12527, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ecin.12527.

Emrath, P., 2021: Government Regulation in the Price of a New Home: 2021, URL https://www.nahb.org/-/media/NAHB/news-and-economics/docs/housing-economics-plus/special-studies/2021/special-study-government-regulation-in-the-price-of-a-new-home-may-2021.pdf.

First Street Foundation, 2020: The First National Flood Risk Assessment. URL https://assets.firststreet.org/uploads/2020/06/first_street_foundation___first_national_flood_risk_assessment.pdf.

First Street Foundation, 2021: The Cost of Climate: America's Growing Flood Risk. URL https://assets.firststreet.org/uploads/2021/02/The_Cost_of_Climate_FSF20210219-1.pdf.

Flavelle, C., 2019: As Wildfires Get Worse, Insurers Pull Back From Riskiest Areas. *New York Times*, URL https://www.nytimes.com/2019/08/20/climate/fire-insurance-renewal.html.

Flavelle, C., 2021: The Cost of Insuring Expensive Waterfront Homes is About to Skyrocket. *The New York Times*.

Flavelle, C., 2022: Why Ian May Push Florida Real Estate Out of Reach for All but the Super Rich. *New York Times*, URL https://www.nytimes.com/2022/10/13/climate/florida-real-estate-hurricane-ian.html.

Frank, T., 2020a: Removing 1 Million Homes from Flood Zones Could Save $1 Trillion. *E&E News*, URL https://www.scientificamerican.com/article/removing-1-million-homes-from-flood-zones-could-save-1-trillion/.

Frank, T., 2020b: Studies Sound Alarm on "Badly Out-of-Date" FEMA Flood Maps. *E&E News*, URL https://www.scientificamerican.com/article/studies-sound-alarm-on-badly-out-of-date-fema-flood-maps/.

Gallagher, J., 2014: Learning about an Infrequent Event: Evidence from Flood Insurance Take-Up in the United States. *American Economic Journal: Applied Economics*, **6 (3)**, 206–233, doi:10.1257/app.6.3.206, URL https://www.aeaweb.org/articles?id=10.1257/app.6.3.206.

Gallagher, R. M., H. Kurban, and J. J. Persky, 2013: Small homes, public schools, and property tax capitalization. *Regional Science and Urban Economics*, **43 (2)**, 422–428, doi:10.1016/j.regsciurbeco.2013.01.001, URL https://linkinghub.elsevier.com/retrieve/pii/S0166046213000021.

Germeshausen, R., 2018: Effects of Attribute-Based Regulation on Technology Adoption - The Case of Feed-In Tariffs for Solar Photovoltaic. *SSRN Electronic Journal*, doi:10.2139/ssrn.3309092, URL https://www.ssrn.com/abstract=3309092.

Gibson, M., and J. T. Mullins, 2020: Climate Risk and Beliefs in New York Floodplains. *Journal of the Association of Environmental and Resource Economists*, **7 (6)**, 1069–1111, doi:10.1086/710240, URL https://www.journals.uchicago.edu/doi/10.1086/710240.

Golnaraghi, M., N. Dufty, and A. Dyer, 2020: Flood Risk Management in Australia., URL http://repo.floodalliance.net/jspui/handle/44111/3872.

Grace, M., R. Klein, and P. Kleindorfer, 2004: Homeowners Insurance with Bundled Catastrophe Coverage. *Journal of Risk and Insurance*, **71 (3)**, URL https://www.jstor.org/stable/3520068.

Harrison, D., G. Smersh, and J. Schwartz, Arthur, 2001: Environmental Determinants of Housing Prices: The Impact of Flood Zone Status. *Journal of Real Estate Research*, **21**, 3–20.

Hino, M., and M. Burke, 2021: The effect of information about climate risk on property values. *Proceedings of the National Academy of Sciences*, **118 (17)**, e2003374118, doi:10.1073/pnas.2003374118, URL https://pnas.org/doi/full/10.1073/pnas.2003374118.

Indaco, A., F. Ortega, and S. Taspinar, 2018: The Effects of Flood Insurance on Housing Markets. 42.

Indaco, A., F. Ortega, and S. Taspinar, 2019: The Effects of Flood Insurance on Housing Markets. *Cityscape*, **21 (2)**, 129–156.

Ito, K., and J. M. Sallee, 2018: The Economics of Attribute-Based Regulation: Theory and Evidence from Fuel Economy Standards. *The Review of Economics and Statistics*, **100 (2)**, 319–336, doi:10.1162/REST_a_00704, URL https://direct.mit.edu/rest/article/100/2/319-336/58457.

Keenan, J. M., and J. T. Bradt, 2020: Underwaterwriting: from theory to empiricism in regional mortgage markets in the U.S. *Climatic Change*, doi:10.1007/s10584-020-02734-1, URL http://link.springer.com/10.1007/s10584-020-02734-1.

Keller, M., M. Rojanasakul, D. Ingold, C. Flavelle, and B. Harris, 2017: Outdated and Unreliable: FEMA's Faulty Flood Maps Put Homeowners at Risk. *Bloomberg*, URL https://www.bloomberg.com/graphics/2017-fema-faulty-flood-maps/.

Kellogg, R., 2020: Output and attribute-based carbon regulation under uncertainty. *Journal of Public Economics*, **190**, 104246, doi:10.1016/j.jpubeco.2020.104246, URL https://linkinghub.elsevier.com/retrieve/pii/S0047272720301109.

Keys, B., 2023: Risky Business: How Climate Change is Changing Insurance Markets, United States Senate Committee on the Budget.

Keys, B., and P. Mulder, 2020: Neglected No More: Housing Markets, Mortgage Lending, and Sea Level Rise. Tech. Rep. w27930, National Bureau of Economic Research, Cambridge, MA, w27930 pp. doi:10.3386/w27930, URL http://www.nber.org/papers/w27930.pdf.

Kousky, C., H. Kunreuther, B. Lingle, and L. Shabman, 2018a: The Emerging Private Residential Flood Insurance Market in the United States. *Risk Management and Decision Processes Center Working Paper*, 53.

Kousky, C., E. O. Michel-Kerjan, and P. A. Raschky, 2018b: Does federal disaster assistance crowd out flood insurance? *Journal of Environmental Economics and Management*, **87**, 150–164, doi:10.1016/j.jeem.2017.05.010, URL http://www.sciencedirect.com/science/article/pii/S0095069617303479.

Kydland, F. E., and E. C. Prescott, 1977: Rules Rather than Discretion: The Inconsistency of Optimal Plans. *Journal of Political Economy*, **85 (3)**, 473–491, doi:10.1086/260580, URL https://www.journals.uchicago.edu/doi/10.1086/260580.

Lee, S., 2022: Adapting to Natural Disasters through Better Information: Evidence from the Home Seller Disclosure Requirement.

Liao, Y., and P. Mulder, 2021: What's at Stake? Understanding the Role of Home Equity in Flood Insurance Demand. *SSRN Electronic Journal*, doi:10.2139/ssrn.3756332, URL https://www.ssrn.com/abstract=3756332.

Lingle, B., and C. Kousky, 2018: Florida's Private Residential Flood Insurance Market. URL https://riskcenter.wharton.upenn.edu/wp-content/uploads/2018/09/Florida-Private-Flood-Issue-Brief.pdf.

Listokin, D., and D. B. Hattis, 2005: Building Codes and Housing. *Cityscape*, **8 (1)**, 21–67.

Livy, M. R., 2018: Intra-school district capitalization of property tax rates. *Journal of Housing Economics*, **41**, 227–236, doi:10.1016/j.jhe.2018.06.008, URL https://linkinghub.elsevier.com/retrieve/pii/S1051137718300342.

Manson, S., Schroeder, Jonathan, D. Van Riper, T. Kugler, and S. Ruggles, 2021: National Historical Geographic Information System: Version 16.0. Minneapolis, MN: IPUMS, URL https://www.nhgis.org/, version Number: 16.0 Type: dataset, doi:10.18128/D050.V16.0.

Morrell, A., 2018: Average new construction lot size remains at record low. *South Florida Agent Magazine*, URL https://southfloridaagentmagazine.com/2018/09/05/average-new-construction-lot-size-remains-record-low/.

Mulder, P., 2021: Mismeasuring Risk: The Welfare Effects of Flood Risk Information, URL https://faculty.wharton.upenn.edu/wp-content/uploads/2017/07/MismeasuringRisk_Mulder2021.pdf.

Mulder, P., 2022: Mismeasuring Risk: The Welfare Effects of Flood Risk Information, URL https://bpb-us-w2.wpmucdn.com/blogs.baylor.edu/dist/a/8176/files/2022/05/S3_Mulder_RiskRatingWelfare.pdf.

National Research Council, 2009: *Mapping the Zone: Improving Flood Map Accuracy*. National Academies Press, Washington, D.C., doi:10.17226/12573, URL http://www.nap.edu/catalog/12573, pages: 12573.

Oh, S., I. Sen, and A.-M. Tenekedjieva, 2021: Pricing of Climate Risk Insurance: Regulatory Frictions and Cross-Subsidies. *SSRN Electronic Journal*, doi:10.2139/ssrn.3762235, URL https://www.ssrn.com/abstract=3762235.

Ostriker, A., and A. Russo, 2022: The Effects of Floodplain Regulation on Housing Markets, URL https://ostriker.github.io/papers/Ostriker-Russo_floodplain-regulations.pdf.

Ouazad, A., and M. E. Kahn, 2022: Mortgage Finance and Climate Change: Securitization Dynamics in the Aftermath of Natural Disasters. *The Review of Financial Studies*, **35 (8)**, 3617–3665, doi:10.1093/rfs/hhab124, URL https://academic.oup.com/rfs/article/35/8/3617/6427560.

Palmon, O., and B. A. Smith, 1998a: A New Approach for Identifying the Parameters of a Tax Capitalization Model. *Journal of Urban Economics*, **44 (2)**, 299–316, doi:10.1006/juec.1997.2072, URL https://linkinghub.elsevier.com/retrieve/pii/S0094119097920726.

Palmon, O., and B. A. Smith, 1998b: New Evidence on Property Tax Capitalization. *Journal of Political Economy*, **106 (5)**, 1099–1111, doi:10.1086/250041, URL https://www.journals.uchicago.edu/doi/10.1086/250041.

Peralta, A., and J. B. Scott, 2019: Moving to Flood Plains: The Unintended Consequences of the National Flood Insurance Program on Population Flows. *Working Paper*, 42, URL https://www.lsu.edu/business/economics/files/microecon-conf-lsu-peralta.pdf.

Pigou, A. C., 1920: *The Economics of Welfare*. Macmillan and Co., Limited, London.

Rouse, C., H. Boushey, and J. Bernstein, 2023: Economic Report of the President. Tech. rep., 273–303 pp. URL https://www.whitehouse.gov/wp-content/uploads/2023/03/ERP-2023.pdf.

Royal, A., and M. Walls, 2019: Flood Risk Perceptions and Insurance Choice: Do Decisions in the Floodplain Reflect Overoptimism? *Risk Analysis*, **39 (5)**, 1088–1104, doi:10.1111/risa.13240, URL https://onlinelibrary.wiley.com/doi/10.1111/risa.13240.

Sastry, P., 2021: Who Bears Flood Risk? Evidence from Mortgage Markets in Florida, URL https://psastry89.github.io/website/psastry_JMP.pdf.

Small, K. A., and H. S. Rosen, 1981: Applied Welfare Economics with Discrete Choice Models. *Econometrica*, **49 (1)**, 105, doi:10.2307/1911129, URL https://www.jstor.org/stable/1911129?origin=crossref.

Song, J., 2021: The Effects of Residential Zoning in U.S. Housing Markets, URL https://jaeheesong.com/s/Jaehee_Song_JMP_share.pdf.

Tavernise, S., and C. Flavelle, 2022: Did Hurricane Ian Bust Florida's Housing Boom? URL https://www.nytimes.com/2022/10/18/podcasts/the-daily/hurricane-ian-floride-housing-market-insurance.html?showTranscript=1.

Train, K. E., 2009: *Discrete Choice Methods with Simulation*. 2nd ed., Cambridge University Press, doi:10.1017/CBO9780511805271, URL https://www.cambridge.org/core/product/identifier/9780511805271/type/book.

U. S. Office of Inspector General, 2017: FEMA Needs to Improve Management of its Flood Mapping Program. URL https://www.oig.dhs.gov/sites/default/files/assets/2017/OIG-17-110-Sep17.pdf.

US Census Bureau, 2022: Historical Population Change Data (1910-2020). URL https://www.census.gov/data/tables/time-series/dec/popchange-data-text.html, section: Government.

Wagner, K., 2021: Adaptation and Adverse Selection in Markets for Natural Disaster Insurance, URL http://www.krhwagner.com/papers/Adaptation%20and%20Adverse%20Selection%20in%20Markets%20for%20Natural%20Disaster%20Insurance%20-%20Katherine%20Wagner.pdf.

Wagner, K. R. H., 2022: Adaptation and Adverse Selection in Markets for Natural Disaster Insurance. *American Economic Journal: Economic Policy*, **14 (3)**, 380–421, doi:10.1257/pol.20200378, URL https://pubs.aeaweb.org/doi/10.1257/pol.20200378.

Wing, O. E. J., P. D. Bates, A. M. Smith, C. C. Sampson, K. A. Johnson, J. Fargione, and P. Morefield, 2018: Estimates of present and future flood risk in the conterminous United States. *Environmental Research Letters*, **13 (3)**, 034 023, doi:10.1088/1748-9326/aaac65, URL https://doi.org/10.1088%2F1748-9326%2Faaac65, publisher: IOP Publishing.

Wing, O. E. J., and Coauthors, 2022: Inequitable patterns of US flood risk in the Anthropocene. *Nature Climate Change*, **12 (2)**, 156–162, doi:10.1038/s41558-021-01265-6, URL https://www.nature.com/articles/s41558-021-01265-6.

Xu, E., C. Webb, and D. Evans, 2019: Wildfire catastrophe models could spark the changes California needs. Milliman White Paper. URL https://www.milliman.com/en/insight/wildfire-catastrophe-models-could-spark-the-changes-california-needs.