Sorting Over Wildfire Hazard

Lala Ma, Matthew Wibbenmeyer, Emily Joiner, Connor Lennon, and Margaret Walls

Working Paper 24-05
May 2024

About the Authors

Lala Ma is a university fellow at Resources for the Future (RFF) and an Associate Professor of Economics and holds the Carl F. Pollard Professorship of Health Economics in the Gatton College of Business and Economics at the University of Kentucky. Her primary research field is environmental economics with a focus on non-market valuation. Her work can be broadly categorized into three areas: valuation as revealed through housing markets, valuation using direct impacts on health or environmental outcomes, and distributional and equity issues related to pollution exposure. Her recent topics of investigation include the environmental health impacts of shale gas development and the housing and equity impacts of flooding and wildfire risk. She received her PhD in Economics from Duke University in 2014.

Matthew Wibbenmeyer is a fellow at RFF. His research studies climate impacts and mitigation within the US land sector, with a special emphasis on wildfire impacts and management. US wildfire activity has accelerated in recent years, leading to increases in property damages, carbon emissions, and health impacts due to smoke. Wibbenmeyer's research studies the impacts of these changes for communities, how these impacts are distributed, and how management choices affect the distribution of impacts. Alongside his work on wildfire, Wibbenmeyer is investigating the role of the US land sector in mitigating climate change, and how policy toward land sector choices may influence the United States' ability to meet climate goals.

Emily Joiner is a research associate at RFF. She works on a suite of topics including carbon dioxide removal, agricultural emissions reduction, and wildfire risk impacts and mitigation. She obtained dual BS degrees in sustainability and economics from Arizona State University and an MS degree in agricultural and resource economics from the University of Arizona (UA). Prior to joining RFF, she held internships with UA's Water Resources Research Center, the Babbitt Center for Land and Water Policy, and the City of Gilbert's Water Department. Joiner's research interests include nonmarket valuation methods, benefit-cost analysis and its regulatory implications, and water economics in the US West.

Connor Lennon is an is economist and data scientist specializing in causal inference and spatial techniques in machine learning and environmental economics. He was an RFF summer intern in 2020.

Margaret Walls is a senior fellow and director of the Climate Risks and Resilience Program, as well as cohost of RFF's podcast, ***Resources Radio***. Walls's research focuses on the impacts of extreme weather, floods, hurricanes, and wildfires on people and communities and the design of programs and policies to equitably enhance resilience to such events. She has written about the benefits of natural infrastructure, such as coastal wetlands and riparian buffers, in reducing flood damages and the evaluation of conservation investment decisions in floodplains and coastal zones. She has also evaluated the role of information and disclosure in driving household location decisions in the context of both wildfire and flood risks. In current research, she is focusing on environmental justice challenges associated with wastewater systems in rural communities facing sea level rise and working in partnership with underserved communities to find ways to invest in natural infrastructure to address coastal flooding. She has also conducted an evaluation of the federal government's implementation of the Justice40 directive, the movement to ensure that 40 percent of the benefits of federal climate and energy investments flow to disadvantaged communities. Walls organized and hosted the six-part environmental justice webinar series at RFF, ***Exposure***, in 2021-2022.

Acknowledgements

This research was supported by the USDA National Institute of Food and Agriculture (NIFA) Agricultureand Food Research Initiative Award No. 2021-67023-34483.

About RFF

Resources for the Future (RFF) is an independent, nonprofit research institution in Washington, DC. Its mission is to improve environmental, energy, and natural resource decisions through impartial economic research and policy engagement. RFF is committed to being the most widely trusted source of research insights and policy solutions leading to a healthy environment and a thriving economy.

Working papers are research materials circulated by their authors for purposes of information and discussion. They have not necessarily undergone formal peer review. The views expressed here are those of the individual authors and may differ from those of other RFF experts, its officers, or its directors.

Sharing Our Work

Sorting Over Wildfire Hazard*

Lala Ma[†], Matthew Wibbenmeyer[‡], Emily Joiner, Connor Lennon, Margaret Walls[§]

May 8, 2024

The costs of natural disasters in the United States have increased in recent years, and, among disaster types, losses to wildfires have risen most sharply. The distribution of costs across households depends, in part, on household incomes and relative willingness to accept (WTA) wildfire risk. Studies have shown that households living in high wildfire hazard areas have higher incomes on average, but this could change with changes in risks, market environments (e.g., insurance), and regulation. In this paper, we use a discrete choice residential sorting model to study relative WTA for wildfire hazard among households with different levels of income and wildfire experience. A spatial discontinuity in California's natural hazard disclosure laws allows us to distinguish aversion to hazard, when it is made salient to buyers at the time of purchase, from preferences toward correlated amenities, such as forest cover. We have three core findings. First, regulatory disclosure matters. In areas where disclosure is required, households are averse to fire hazard. Second, aversion is increasing in income, which suggests that lower-income households may replace high-income households in high-hazard areas, raising concerns about distributional justice. Finally, individual experience with wildfire events does not increase aversion to hazard, suggesting that experience is not a replacement for disclosure in motivating informed decision-making.

Classification: Social Sciences; **Keywords:** Natural Disasters; Wildfire; Sorting; Information Disclosure

*This research was supported by the USDA National Institute of Food and Agriculture (NIFA) Agriculture and Food Research Initiative Award No. 2021-67023-34483.
[†]University of Kentucky, Department of Economics, Lexington, KY. Email: lala.ma@uky.edu
[‡]Resources for the Future, Washington, DC. Email: wibbenmeyer@rff.org.
[§]Resources for the Future, Washington, DC. Email: walls@rff.org.

1 Introduction

Costs of natural disasters in the US are increasing, and insured losses from disasters has topped $70 billion in each year since 2020 (Insurance Information Institute, 2023). The growing exposure of people and property to natural hazards, such as floods and wildfires, is a significant contributor to this trend. Wildfire is the peril with the fastest-growing damages (Swiss Re Institute, 2023), and homes in the wildland-urban interface (WUI), the area where development intermingles with wildlands, account for most of the properties lost (Kramer et al., 2019). By some accounts, WUI is the fastest-growing land use type in the conterminous United States (Radeloff et al., 2018), and more than 600,000 new homes are expected to be built in high-wildfire-hazard areas in California alone by 2050 (Mann et al., 2014).

Household demand to live in high-hazard areas is determined by preferences and income, the set of available options, and information about those choices. Households may choose such areas because they are relatively lower cost or because they prefer rural amenities that contribute to fire hazard, such as forest cover. Evidence indicates that despite significant heterogeneity in both income and property values within high-fire-hazard areas, the value of properties located in them are, on average, higher (Wibbenmeyer & Robertson, 2022). This suggests that households in these areas are more likely to have been "pulled" there by desirable amenities correlated with hazard than "pushed" there by high costs-of-living elsewhere (although explanations likely vary across locations).

However, the balance of factors affecting the choice to live in high-fire-hazard areas is changing rapidly. Along with increasing damages from wildfire events (Higuera et al., 2023), homeowner's insurance in these areas has become more expensive and more challenging to obtain, especially in California (Cignarale et al., 2017; Liao et al., 2022). Homeowners are increasingly expected, and sometimes required, to maintain "defensible space" around their homes, which may be costly and/or reduce the value of amenities associated with these areas. Public perceptions about the risks of wildfire hazards are also evolving due to recent exposure to wildfire events and policies that regulate information provision. These changes could influence how households sort across the landscape with respect to wildfire hazard, the distributional incidence of wildfires, and which groups predominantly benefit from implicit subsidies to live in wildfire hazard areas (Baylis & Boomhower, 2023).

We explore how the changing wildfire risk environment will influence sorting over wildfire hazard. Specifically, we apply a discrete choice residential sorting model to examine how willingness to accept (WTA) wildfire hazard varies among households with different levels of information about it, different levels of experience with wildfires, and varying incomes. Our empirical analysis focuses on California in 2009–2018. California makes for a desirable study area for three reasons. First, it has been especially hard hit by rising wildfire damages; of the top 10 fires in US history with respect to insured losses, eight have occurred in California, seven since 2017.[1] Second, as mentioned, increased risk across the state has presented new obstacles to living in wildfire hazard areas, including challenges obtaining insurance. Third, its regulatory laws regarding hazard disclosure provide a unique opportunity to identify WTA (e.g. Ma et al., 2024; Garnache, 2023).

Our study confronts three primary empirical challenges. First, standard revealed preferences models assume that agents are perfectly informed about the good of interest (Rosen, 1974), which may bias demand estimates if that assumption is violated (Ma, 2019). If buyers are uninformed, they may be less averse to purchasing homes with high degrees of hazard. Second, estimation of hedonic and sorting equilibrium models often suffers from omitted variables bias (Greenstone, 2017). We cannot hope to control for all factors that may be correlated with hazard, potentially including tree cover, proximity to recreational opportunities, and distance from population centers and places of work. Conflating these correlated amenities (or disamenities) with hazard could bias estimates of aversion toward hazard. Finally, we face the standard challenge in the empirical household sorting literature that home prices are endogenous.

We address the first two challenges using a boundary discontinuity design in conjunction with a difference-in-differences framework. The boundary discontinuity design exploits a spatial discontinuity in California wildfire hazard disclosure laws: buyers of homes in designated areas are required to be notified about hazard at the time of purchase. Recovering the relative prices of homes sold in disclosure-regulated areas ensures that estimated preferences

[1]This is based on data from the Insurance Information Institute (`https://www.iii.org/fact-statistic/facts-statistics-wildfires`) and estimated insured losses of $3.2 billion from the 2023 wildfire in Lahaina, Hawaii.

are based on choices made under full information and explicitly addresses the first identification challenge—that identifying WTA for hazard requires that households be informed of it. To ensure that estimated differences in preferences are due to wildfire hazard—and not to preferences over underlying differences between disclosure and nondisclosure areas that correlate with it, the second of our identification challenges—we additionally contrast preferences for properties that are nearby to one another but differ in disclosure requirements. The identifying assumption is that unobserved variables correlated with wildfire hazard should not differ substantially for properties near the same boundary, even though the disclosure requirement does.

To address price endogeneity, the third identification challenge, we follow Bayer et al. (2016) in adopting a strategy based on household moving costs. The basic intuition is that the shadow value of income can be estimated based on moving costs, which are related to the value of the previous home and arguably exogenous to unobserved characteristics of the new home. This is because moving costs are partially determined by Realtor fees, which are based on the value of the home sold. Empirically, we track buyer locations over time with a name-matching algorithm and use transaction values from the sales of a previous home to measure moving costs.

The ability to track home buyers over time allows us to investigate another possible determinant of household location choice: experience with wildfires. In the hedonic literature on flooding, a strong price effect has been found after a flood at a given home's location (e.g. Hallstrom & Smith, 2005; Kousky, 2010). Studies have attributed these findings to an array of behavioral factors that affect decision making, such as salience, reliance on heuristics, and recency bias. We take this a step further by considering the effect of experience with wildfires at *previous* locations at which a household has lived on aversion toward wildfire hazard.

This paper contributes to several strands of literature. The first is the extensive hedonic valuation literature evaluating household WTA risk from natural hazards. Much of this literature focuses on flooding and hurricanes. Studies from these contexts have found that risk information, including flood risk maps and disclosure requirements (Gibson & Mullins, 2020; Hino & Burke, 2021) and the information conveyed by events themselves (e.g. Hallstrom & Smith, 2005; Kousky, 2010; Bin & Landry, 2013; Zivin, Liao, & Panassie, 2023), is important in shaping household market behavior. These results are consistent with findings in the wildfire context, where a more limited set of studies has found that recent fires, hazard maps, and disclosure laws can each decrease home prices in high-hazard areas (Donovan, Champ, & Butry, 2007; McCoy & Walsh, 2018; Garnache, 2023; Ma et al., 2024). The literature that models sorting over natural hazards is smaller but includes a few studies that use discrete choice residential sorting models to evaluate heterogeneous aversion toward flood risk and extreme heat (Fan & Davlasheridze, 2015; Fan, Klaiber, & Fisher-Vanden, 2016; Fan, Fisher-Vanden, & Klaiber, 2018; Bakkensen & Ma, 2020; Hyde, 2023). Our study innovates on this line of work on a few dimensions. First, our model allows for moving costs, which has been shown to bias willingness to pay (WTP) estimates (Bayer, Keohane, & Timmins, 2009). Second, the strategy to apply moving costs to identify the marginal utility of income avoids an instrumental variables approach, which frequently relies on arguably tenuous exogeneity assumptions. Third, our study incorporates the effect of experience with disasters on location choice, enabled by linking homebuyer decisions over time. That experience has been shown to affect consumer choice in a wide range of consumer markets (Ackerberg, 2003; Chernew, Gowrisankaran, & Scanlon, 2008; Ashworth et al., 2021; Gallagher, 2014). The literature on the effects of individual experience in climate adaption is thin. Given the potentially staggering effects of natural disasters, prior experience may be an important determinant in the choice of wildfire hazard exposure.

This study also relates to a broader body of literature on wildfire and the WUI. A significant portion of it is focused on understanding how the WUI is expanding due to real estate development and how this is likely to shape wildfire risk (e.g. Radeloff et al., 2018; Mann et al., 2014). In contrast, we investigate factors that affect household decisions over whether to live in high-hazard areas, taking the housing stock as given, which is better positioned to study distributional changes in exposure to wildfire hazard rather than changes in exposure to wildfire risk overall. In this way, our paper relates to Issler et al. (2020), which examines environmental gentrification after wildfire events, but we study heterogeneous sorting over *hazard* rather than induced by previous fire events.

2 Wildfire Hazard Disclosure Laws in California

Home buyers in California may receive more or less information about a home's wildfire risk depending on its location. Since 1998, California law has required sellers of homes in designated wildfire hazard areas to disclose hazard to buyers (California Legislature, 1997) through a Natural Hazard Disclosure Statement (NHD), which indicates that the property lies within a designated wildfire hazard area.[2] The NHD is provided before escrow, at which point, buyers have three days to back out of the sale.[3]

The wildfire hazard disclosure rules that apply are determined by properties' locations relative to both responsibility areas and fire hazard severity zones (FHSZs). Responsibility areas establish who is financially responsible for fires and are divided into local (LRA), state (SRA), and federal (FRA). These correspond approximately to incorporated lands, unincorporated private lands, and federal lands, respectively.[4] Since 1981, the California Department of Forestry and Fire Protection (CAL FIRE) has mapped fire hazard within SRAs based on fuels, topography, fire weather, and potential for high wind events that can fuel large wildfires. The FHSZs are divided into moderate, high, and very high within the SRA. In 1992, CAL FIRE also began to map very-high FHSZs within LRAs, although adoption of these zones was left to local governments.[5] California requires disclosure for any homes sold within moderate, high, or very-high FHSZs within SRAs. Within LRAs, however, disclosure is required only for homes in very-high FHSZs.

Areas in California requiring wildfire hazard disclosure also require compliance with building codes intended to reduce vulnerability to fire (Baylis & Boomhower, 2022). These codes place restrictions on building materials that can be used in newly constructed homes or exist-

[2] Disclosure requirements also apply to Special Flood Hazard Areas, earthquake fault zones, and seismic hazard zones.

[3] NHDs do not inform buyers about the relative level of wildfire hazard within an area, only that the property lies within a designated hazard area. Although households cannot easily observe the *level* of wildfire hazard, they do observe other property characteristics, such as nearby wildlands and vegetation, and topography, known to influence wildfire hazard. Our hypothesis is that disclosure may increase the salience of these features and their effects on hazard.

[4] One reason that responsibility areas may not precisely correspond to these boundaries is the "balance of acres" arrangements, through which jurisdictions trade fire responsibilities in some areas to maximize efficiency (Starrs et al., 2018).

[5] According to data collected by Miller, Field, & Mach (2020), most communities in California have adopted very-high FHSZs.

ing homes at the time of a major update, such as a roof replacement. Use of building codes to manage wildfire risk in California dates to 1994; however, they were substantially strengthened in 2008. Before 2008, wildfire building codes pertained mainly to roofing materials, but after 2008, new homes built in high-hazard areas (moderate, high, or very-high FHSZs in SRAs or very-high FHSZs in LRAs) had to have fire-resistant eaves, siding, windows and doors, and vents, among other requirements.[6]

3 Data

3.1 Property Transaction Data

To estimate the model of housing location choice, we use transaction and assessors' data from Zillow's Transaction and Assessment Dataset (ZTRAX).[7] We limit ZTRAX sales transactions to the sample of home purchases in California 2009–2018. For each one, we link transactions data, which include sales price, with assessors' data from ZTRAX to provide property-level covariates, including lot size, number of bedrooms, square footage, and year built. We restrict our sample to arms-length purchases of single-family homes. After these sample restrictions, our data set contains 2.8 million transaction observations.

We identify buyer demographic information in each sale by merging transactions from ZTRAX to Home Mortgage Disclosure Act (HMDA) Loan Application Register (LAR) data, following a procedure outlined by Bayer et al. (2016) and applied to ZTRAX data by Billings (2019). The LAR is a collection of data on demographics for individuals covering all property-backed loans in the United States.[8] The dataset is publicly available and contains loan-level records identifying the year, property census tract, loan amount, and an alphanumeric lender identifier. Each record also contains data on lendee demographics, including race and household income. Because high-fire-hazard areas in California are relatively homogeneous

[6]To ensure that we are identifying WTA for wildfire hazard rather than preferences for fire-resistant home characteristics, we test the robustness of our results to dropping homes built after 2008.

[7]Data are provided by Zillow through ZTRAX. More information on accessing the data can be found at `http://www.zillow.com/ztrax`. The results and opinions are our own and do not reflect the position of Zillow Group.

[8]This dataset was initially proposed to allow public scrutiny of lending decisions and expose potential discrimination.

in race, we focus on sorting based on household income. We merge transactions data from ZTRAX to HMDA loan data by identifying unique matches on year, census tract, loan amount, and lender name. Because HMDA data contain alphanumeric lender identifiers but not names, we first identify unique matches on only year, census tract, and loan amount. We then use these matches to construct a dictionary that translates lender identifiers to names. This represents an advance on the ZTRAX–HMDA merge procedure described by Billings (2019)—which we otherwise follow closely—and enables us to match demographics to a greater number of ZTRAX transactions than Billings (2019) do for California in the same period. More information on the matching procedure and its performance is provided in Appendix C.

Following Bayer et al. (2016), we estimate moving cost as 6 percent of the value of the home from which a household moved. We identify each buyer's most recent home (i.e., the "origin" home) using an algorithm that identifies unique matches between the buyer's name and names of sellers listed for transactions around the same time. The algorithm first identifies unique matches on first and last name, then considers matches on first initial and last name. Finally, in a third step, it identifies matches using a phonetic algorithm to allow for misspellings. Due to the matching procedure, our final data set contains financial moving costs only for households that moved within California and not those that moved into the study area from another region. We are also not able to identify moving costs for first-time buyers. Nevertheless, we are able to identify previous neighborhoods for 39 percent of the transactions in our merged ZTRAX–HMDA data set. After restricting the data set to only those observations, we are able to match to HMDA loan records and previous home sales, our final data set contains $256,821$ transactions within California for 2009–2019. For more information about our procedure for identifying previous neighborhoods, see Appendix D.

Although the rate at which we match ZTRAX transactions observations to HMDA records compares favorably to other studies (see appendix C.2), and we identify origin homes at a similar rate, one might nevertheless be concerned about potential bias caused by limiting the sample to only those transactions that successfully pass through these two matching processes. Table 1 provides demographic summary statistics for homeowners in California (based on the American Community Survey from the US Census), alongside demographic summary statistics from the raw HMDA data, summary statistics for the set of HMDA-matched ZTRAX transactions, and transactions for which we also identify the origin home. Demographics of homeowners, as measured by the US Census, differ slightly from demographics in the set of HMDA records we use to match transactions. Most notably, the share of Asian residents is more than 4 percentage points higher in the HMDA data, and median household income is more than $10,000 higher. However, these differences may be due to differences in how race and income are measured. Demographic differences are largely minor across Columns 2, 3, and 4 (although we do match a disproportionate number of Hispanic buyers to HMDA loan records). Likewise, the demographics of HMDA-matched buyers also matched to their origin home (Column 4) are similar to those in the full set of HMDA-matched buyers. Altogether, the statistics indicate that the HMDA and origin home matching processes do not appear to substantially alter the demographics of our sample.

We supplement property-level variables from the ZTRAX data set with a large suite of property- and neighborhood-specific variables from other sources. We measure each property's distance to the nearest protected area[9] and whether it is within an incorporated area.[10] The latter variable is important because LRAs are significantly more likely to be incorporated than SRAs. Measuring incorporated status allows us to control for benefits or costs of living in an incorporated area that are correlated with whether a property is in an SRA—and thus whether it faces disclosure requirements, which is more likely for SRA properties. We obtain the share of White non-Hispanic residents in each property's block group from the 2010 US Census Summary File 1 and average standardized test scores from the California Department of Education. We calculate the latter from the average of the percentage of students in each property's district who met or exceeded math and language arts testing standards. Finally, we also map properties to various wildfire-related variables of interest, including historical fires (to infer fire experience) and FHSZs and responsibility areas, which jointly determine information disclosure requirements. We return to these sources of data in the next subsection.

[9]Protected areas are based on data from the USGS Protected Areas Database. We measure the distance to the nearest protected area classified as GAP Status 1 or 2, which are permanently protected areas on which extractive uses are not allowed.

[10]Incorporated area boundaries are available from CAL FIRE.

Table 2 summarizes property- and neighborhood-specific variables. According to Column 1, the average home in our sample was approximately three bedrooms and 2,000 square feet, about 40 years old at the time of sale, and sold for $612,000. On average, properties were approximately 15 km from the nearest protected area, and the share of White non-Hispanic residents in their census tracts was 50.1 percent. Column 2 describes the characteristics of properties in our sample in areas where wildfire hazard disclosure is required; Column 3 describes characteristics of properties in areas where it is not. As noted in Section 2, disclosure is required in moderate, high, or very-high FHSZs within SRAs and very-high FHSZs within LRAs. These areas tend to be more rural than areas with no disclosure requirements. Correspondingly, homes in areas with disclosure requirements are, on average, larger (greater square footage), on larger lots, closer to protected areas, and less likely to be within an incorporated area. Homes in areas requiring disclosure are more expensive on average, consistent with findings in Wibbenmeyer & Robertson (2022). They are also newer, consistent with research on the expanding WUI (Mann et al., 2014; Radeloff et al., 2018), and tend to be in areas with higher shares of White residents. Last, although the next section discusses the wildfire hazard potential (WHP) and previous fire measures in more detail, homes in areas requiring disclosure have higher hazard and are more likely to have been near a wildfire in the five years before sale. Differences between characteristics of homes in areas requiring and not requiring disclosure highlight the need for our boundary discontinuity approach, which compares preferences for homes very close to one another but on either side of the boundary dividing those areas.

3.2 Disclosure Requirement, Wildfire Hazard, and Wildfire Data

As described in Section 2, wildfire hazard disclosure requirements in California are determined by the responsibility area and designated FHSZ classification. We use the "nine-class" FHSZ data from CAL FIRE, which include moderate, high, and very-high FHSZs within the SRAs in 2007, and very-high FHSZs that were adopted by communities within LRAs in the ensuing years. In addition, based on CAL FIRE fire hazard modeling, remaining LRAs are divided into moderate and high and FRAs into moderate, high, or very high. This results in nine responsibility area-by-fire hazard classes.[11] We use these data to identify areas where disclosure requirements apply during the period of our data.

Our boundary discontinuity design infers preferences based on household choices on either side of the disclosure boundary. To control for fixed differences in neighborhood quality across that boundary, we first identify boundaries between disclosure and nondisclosure areas in our study area. We then divide these into discrete segments, following a procedure described in Bakkensen & Ma (2020) and Ma et al. (2024), and match each property to the nearest boundary segment. Within our study area, the procedure results in 51,371 segments, with an average length of 3.4 km.

We are interested in differences in aversion to wildfire hazard across households based on varying information about hazard due to differences in disclosure regulation at the time of purchase. Wildfire hazard varies significantly both within and across areas with and without disclosure requirements. In part, this is because whether a home faces a disclosure requirement is a function of not merely hazard but also its responsibility area. However, even areas judged by CAL FIRE to contain similar hazard appear to have significant variation. We measure wildfire hazard at each property's location using Wildfire Hazard Potential (WHP), a national wall-to-wall measure of the relative potential for fires that would be difficult to contain at any given location on the landscape (Dillon, Menakis, & Fay, 2014). In contrast to CAL FIRE's fire hazard maps, which provide contiguous hazard zones for three discrete levels, WHP provides a continuous ordinal measure at a given location. Moreover, it is calculated based on a different, and arguably more comprehensive, approach, which integrates spatial fire occurrence data, fire spread modeling (which takes as inputs fuels and topography data, as CAL FIRE also uses), and a set of weights representing fire's resistance to control across different fuel types. The resulting measure ranges from 0 to approximately 100,000; however, the distribution of WHP across the landscape is highly skewed. In our sample of homes, 95 percent of observations have WHP values below 732.

Figure 1 shows substantial overlap in the distribution of WHP across hazard classes; although wildfire hazard is on average higher in areas where disclosure is required—as one would expect—WHP nevertheless has considerable variation within areas where disclosure

[11] The data also include a "None" category indicating areas rated as having low or no fire hazard.

is required and areas where it is not. This point is further underlined by Figure 2, which shows the boundaries of disclosure requirement areas (stippled) overlaid against a map of WHP. Panel A shows variation in WHP across California. Panel B provides an inset map of Los Angeles County, with a clearer view of the variation in wildfire hazard within and across disclosure requirement boundaries. The boundaries do not correspond precisely to discontinuous changes in wildfire hazard. We show later that close to the boundary between disclosure and nondisclosure areas, hazard is similar on either side of the boundary.

In addition to fire hazard, we use historical fire perimeters to measure household experience with fires at the previous home and their amenity effects. We obtain large fire perimeters from the USGS Monitoring Trends in Burn Severity data set (Eidenshink et al., 2007). Identifying the home a household moves from, which we do primarily to identify moving costs, also allows us to measure whether it experienced nearby fires at its previous home. When we observe the date at which the household purchased the origin home, we define fire experience using a binary variable indicating whether a fire occurred within 10 km during the time it lived at its origin home;[12] when the purchase date of the origin home is not observed, we assume the household lived there since the start of the sample period (2009). To control for amenity effects of nearby fires at chosen homes, we also generate variables indicating whether properties have been inside the perimeter of a large fire in the previous five years. A map of fires in California over the period of our study is provided in Figure A.1. Fires are frequent in areas with high wildfire hazard, as mapped in Figure 2. The figure also displays a noticeable increase in burned area in recent years.

Table 3 summarizes selected home characteristics, including wildfire hazard characteristics, by buyer characteristics. The first three columns split the sample into terciles of household income and summarize characteristics of homes purchased by buyers in each tercile. As expected, higher-income buyers purchase higher-value, larger, and newer homes on average. They also tend to purchase homes in areas with a higher share of White residents (likely reflective of correlated amenities). Interestingly, wildfire hazard is highest for homes

[12]We expect that the effect of fire experience on preferences is decreasing with distance to fire. The 10 km threshold was chosen based on the price impacts documented in hedonic studies and balancing the number of households with fire experience (which increases with the distance threshold) against diluting its average effect.

purchased by buyers in the middle tercile; it is similar for low- and high-income households. The final two columns of the table compare characteristics of homes purchased by buyers who have not and have, respectively, experienced a nearby fire at their most recent home. Home characteristics are similar across the two groups, except with respect to variables related to fire hazard and disclosure. On average, buyers who experienced a nearby fire choose homes in higher-hazard areas.

4 Theory & Estimation

4.1 Model Setup

A household i, described by K characteristics z_k^i, in period t makes a residential location choice, d_t^i, based on its preferences for location characteristics and the costs of moving. If it chooses to move, then it incurs a financial moving cost FMC_t^i, equal to 6 percent of the value of the house it occupies, and then chooses one of J neighborhoods to move to, $d_t^i \in \{1, \ldots, J\}$. If it chooses to stay, then no costs are incurred, and $d_t^i = 0$.

Characteristics of location choices are described by X_{1jt}, ξ_{jt}, and X_{2jt}. Preferences for X_{1jt} and ξ_{jt} do not vary by individual attributes z_k^i, whereas preferences for X_{2jt} do. Choice attributes in X_{1jt} broadly represent house structural characteristics that are observed by the household and the analyst; ξ_{jt} represents unobserved (to the analyst) neighborhood characteristics for which households have preferences. X_{2jt} includes attributes that describe wildfire hazard, the information buyers of homes at a given location have about it, and neighborhood characteristics that may be correlated with it. All households choosing to live in neighborhood j at time t pay a flow cost of housing given by C_{jt}.[13] Finally, households may have idiosyncratic preferences over specific neighborhoods, which are not observed by

[13] As we only observe sales prices in years during which a given home sold, we interpolate prices in other years based on observed prices and average year-over-year price changes for other homes. First, we use homes with repeated sales and linearly interpolate log prices for intervening years between sales. We then use observed and interpolated prices and calculate the median annual percentage change in price amount within Southern (Imperial, Inyo, Kern, Los Angeles, Mono, Orange, Riverside, San Bernardino, San Diego, San Luis Obispo, Santa Barbara, Tulare, and Ventura counties) and Northern California (all other counties). Finally, we average regional median percentage changes within each two-year period, 2009–2018, and use these average changes to interpolate missing prices for periods in which transactions were not observed and prices could not be interpolated based on repeated sales.

the analyst. We denote these unobserved preferences ϵ^i_{jt}.

Neighborhoods within our data set are defined as unique combinations of county, ZIP code, disclosure requirement (yes/no), FHSZ classification (none, moderate, high, very high), and an ID associated with the nearest regulated area boundary. Together, combinations of these variables define ~2,584 distinct neighborhoods within 1 km of regulatory boundaries, which comprise the set of choices over which households may sort.[14] We assume that each decision period is two years, during which a household decides whether to move from its current location or remain. Our main sample includes 23,979 households choosing neighborhoods over each of the five two-year periods from 2009 to 2018.

A household i's indirect utility of choosing location j at time t is given by

$$V^i_{jt} = \alpha_{x_1} X_{1jt} + \alpha^i_{x_2} X_{2jt} - \alpha_C C_{jt} - \beta_{FMC} 1\left[d^i_t \neq 0\right] \cdot FMC^i_t + \xi_{jt} + \epsilon^i_{jt}, \tag{1}$$

where

$$\alpha^i_{x_2} = \alpha_{0,x_2} + \sum_{k=1}^{K-1} \alpha_{k,x_2} z^i_k. \tag{2}$$

In practice, we consider heterogeneous preferences for the variables in X_{2jt} over two household characteristics: income and, in some specifications, household experience with wildfire. Income is mapped to one of three terciles ($K = 3$ household income types). Household fire experience is a binary variable indicating a wildfire within 10 km of its previous location during its tenure.[15] The vector of neighborhood characteristics X_{1jt} includes number of bedrooms, age of the house, log(housing square feet), number of fires within 5 years, and an indicator for whether the neighborhood is in an incorporated area. Neighborhood characteristics X_{2jt}, for which household preferences vary by individual type (e.g., income), include whether the neighborhood regulates wildfire disclosure, log(wildfire hazard potential), log(lot size), average test score, share of the neighborhood that is White, log(distance to the nearest protected area), and number of fires within the last five years.

[14] We base the total number of available choice locations in each period based on choices that we observe to have positive shares. Thus, the total number of choice–period observations is smaller than $J \times 5$ because we do not observe selections of some neighborhoods in all five two-year periods.

[15] When we do not have the date that the household moved to its previous location, we assume that it has lived there since the beginning of our sample period.

The household's problem is static:[16] Household i will choose alternative $d_t^i = j$ if it yields the highest flow utility among all other alternatives:

$$d_t^i = j \text{ if } V_{jt}^i \geq V_{j't}^i \;\; \forall \;\; j' \neq j \tag{3}$$

We assume that household idiosyncratic tastes for choices are distributed i.i.d. Type I Extreme Value. Therefore, the expected probability that a household chooses residence j has a closed-form expression (McFadden, 1978):

$$P_{jt}^i \equiv Pr\left(V_{jt}^i \geq V_{j't}^i \;\; \forall \;\; j' \neq j \mid X, C, Z, FMC\right) = \frac{e^{V_{jt}^i}}{\sum_{j'=0}^{J} e^{V_{j't}^i}}. \tag{4}$$

Section 4.2 discusses our strategy for estimating Equation 4, which is challenging due to the large number of choices in the choice set.

4.2 Estimation

To facilitate estimation, we separate the indirect utility V_{jt}^i into choice- and individual-specific components:

$$V_{jt}^i = \delta_{jt} + \left(\sum_{k=1}^{K-1} \alpha_{k,x_2} z_k^i\right) X_{2jt} + \beta_{FMC} 1\left[d_t^i \neq J+1\right] \cdot FMC_t^i + \epsilon_{jt}^i, \tag{5}$$

where

$$\delta_{jt} = \alpha_{x_1} X_{1jt} + \alpha_{0,x_2} X_{2jt} - \alpha_C C_{jt} + \xi_{jt}. \tag{6}$$

The choice-specific component, δ_{jt}, represents the mean utility of the base (or omitted) group, which consists of those in the lowest-income tercile.[17] The remaining components not absorbed in δ_{jt} include the heterogeneous taste parameters (α_{k,x_2}) and the financial disutility of moving (β_{FMC}). The parameter α_{k,x_2} is the coefficient on the interaction between the

[16] Our analysis focuses on the decision of whether to move in a given period (and if so, where to move to) rather than the decision of when to move. The literature has shown that assuming individuals are myopic in their location choices can bias preference estimates (Bayer et al., 2016; Bishop & Murphy, 2011, 2019; Ma, 2019).

[17] When we additionally allow preferences to vary by fire experience, the base group becomes those in the lowest-income tercile *without* any fire experience.

individual's type and a neighborhood attribute in X_{2jt}. For any one of these attributes, this coefficient represents the additional utility that a household of type k receives *relative to* the base group. For example, for wildfire hazard potential, the coefficient α_{2,x_2} represents the additional utility (or disutility) households of increases in wildfire hazard for households in income tercile 2 receive relative to households in income tercile 1.

Estimation of V^i_{jt} proceeds in two stages, following Berry, Levinsohn, & Pakes (1995). Stage 1 uses maximum likelihood estimation (MLE) to recover heterogeneous preference parameters, the moving cost parameter, and mean utilities. Stage 2 is a regression that decomposes the mean utility estimates from Stage 1 to recover the remaining base group parameters (or the "mean utility decomposition"). Details for each stage are described next.

Stage 1 In the first stage, we use MLE to estimate the J mean utility parameters for each period (δ_{jt}) and the individual and choice-specific parameters ($\alpha_{k,x2}, \beta_{FMC}$). To reduce computational burden, we reformulate our problem as a nested logit, in which the choice to move or stay comprises an upper nest, and the choice of where to move (conditional on moving) is a lower nest. This allows us to split estimation of the first-stage parameters into two steps: (1) preference parameters $\alpha_{k,x2}$ and δ_{jt} based on the decision of where to move; and (2) the financial moving cost parameter based on the decision of whether to move. Although this is not efficient, nesting the problem in this way significantly reduces computation time.

For the lower nest, we estimate the mean utilities and heterogeneous preference parameters using the sample of movers only. Following Berry (1994), we use a contraction mapping routine to recover δ_{jt} for each guess of $\alpha_{k,x2}$, where we normalize the mean utility of one choice in each period to be 0. Once we recover estimates of δ_{jt} and $\alpha_{k,x2}$, we can use them to form the inclusive value of moving in the moving decision. That value represents the utility a household receives from its choice among neighborhoods, conditional on moving. Combining the inclusive value with utility cost of financial moving costs, the utility of moving (i.e., $d^i_t \neq 0$) is the following:[18]

$$V^i_{d^i_t \neq 0,t} = \log\left(\sum_{j'=1}^{J} e^{V^i_{j't}}\right) + \beta_{FMC} FMC^i_t. \tag{7}$$

[18]We assume that the nesting parameter is 1 so that this problem is equivalent to a multinomial logit.

The utility of staying is as before (i.e., $V^i_{0,t}$). Combining these definitions of utility for movers and nonmovers allows us to estimate the financial moving cost parameter based on observed moving decisions using a binary logit. This stage returns the moving cost parameter, β_{FMC}.

Stage 2 The second stage regresses mean utility estimates on neighborhood attributes to recover the preferences for these attributes:

$$\hat{\delta}_{jt} = \alpha_{x1} X_{1jt} + \alpha_{0,x2} X_{2jt} - \alpha_C C_{jt} + \xi_{jt}. \tag{8}$$

The coefficients on attributes for which households have heterogeneous preferences ($\alpha_{0,x2}$) represent the preferences of the base group, and those on the remaining attributes (α_{x1}, α_P) represent the average preferences of all households. We can introduce other spatial or temporal fixed effects at this step.

4.3 Identification

As explained in Section 1, we are concerned about three main issues in estimating equation 8: (1) endogeneity of price (2) endogeneity of wildfire hazard, and (3) incomplete information about hazard. First, price C_{jt} is likely correlated with unobserved neighborhood quality, ξ_{jt}. Rather than estimate α_C using OLS, we assume that equals the negative of the moving cost parameter β_{FMC}, which also measures the marginal utility of income. We then substitute α_C with β_{FMC} and move cost of living ($\beta_{FMC} C_{jt}$) to the left-hand side of equation 8, so that the dependent variable now becomes a price-adjusted mean utility. Finally, we convert the dependent variable into dollars by dividing it by -β_{FMC} so that coefficients are directly interpreted as a dollar value of impact. The following gives the updated estimation equation:

$$\frac{\hat{\delta}_{jt} - \hat{\beta}_{FMC} C_{jt}}{-\hat{\beta}_{FMC}} = \alpha_{x1} X_{1jt} + \alpha_{0,x2} X_{2jt} + \xi_{jt}, \tag{9}$$

where the parameters with $\widehat{\ }$s are estimated in the first stage. The advantage of this strategy is that our measure of moving cost is exogenous to unobserved characteristics of the chosen location d^i_t because moving cost is a function of the value of household i's previous home.

Next, home buyers may not be fully informed about hazard in potential locations. We address this challenge by using California hazard disclosure laws. We compare aversion to wildfire hazard among buyers in areas where disclosure is required to that in areas where it is not. In disclosure areas, buyers have more information about hazard; therefore, we expect increases in hazard to reduce WTP by a greater amount than it would outside of disclosure areas. Crucially, although disclosure rules are a function of FHSZs, hazard varies substantially within these zones; therefore, it is possible to compare household preferences for homes facing similar hazard but differing with respect to disclosure. We separate out the variables of interest related to wildfires in the mean utility decomposition:

$$\begin{aligned}\frac{\hat{\delta}_{jt} - \hat{\beta}_{FMC} C_{jt}}{-\hat{\beta}_{FMC}} &= \alpha_{0,r} \log(1 + WHP_j) + \alpha_{0,reg} Disc_j + \alpha_{0,DD} Disc_j \cdot \log(1 + WHP_j) \\ &+ \alpha_{x1} X_{1jt} + \alpha_{0,x2} X_{2jt}^{-wildfire} + \xi_{jt},\end{aligned} \tag{10}$$

where $log(1+WHP_j)$ is the log of wildfire hazard potential, $Disc_j$ is an indicator for whether the choice requires disclosure, and $X_{2jt}^{-wildfire}$ are the remaining nonwildfire variables in X_{2jt}. In addition to distinguishing choices where disclosure is required, we also differentiate *individuals* based on their fire experience, proxied by whether a wildfire occurred within 10 km of their previous residence during their tenure there.

Finally, wildfire hazard is often correlated with amenities, some of which may be unobserved. To mitigate confounding of preferences for amenities that are correlated with wildfire risk when comparing locations far apart, we additionally control for nearest-disclosure-boundary fixed effects, BD_j, in Equation 10 and limit our analysis to locations within 1 km of disclosure boundaries.[19] Our final estimating equation for Stage 2 is the following:

$$\begin{aligned}\frac{\hat{\delta}_{jt} - \hat{\beta}_{FMC} C_{jt}}{-\hat{\beta}_{FMC}} &= \alpha_{0,r} \log(1 + WHP_j) + \alpha_{0,reg} Disc_j + \alpha_{0,DD} Disc_j \cdot \log(1 + WHP_j) \\ &+ \alpha_{x1} X_{1jt} + \alpha_{0,x2} X_{2jt}^{-wildfire} + BD_j + \xi_{jt}.\end{aligned} \tag{11}$$

Consistent with this specification, in the first stage, we estimate $\hat{\delta}$, β_{FMC}, and the individual-

[19]For robustness, we re-estimate our model using the full sample but separating out location choices within and outside of 1 km of boundaries and modifying the specification to one of triple differences.

specific parameters $\alpha_{k,x2}$ according to

$$V_{jt}^{i} = \delta_{jt} + \sum_{k=1}^{K-1} \Big(\alpha_{k,r} \log(1 + WHP_j) + \alpha_{k,disc} Disc_j + \alpha_{k,DD} Disc_j \cdot \log(1 + WHP_j) + \alpha_{k,x2} X_{2jt}^{-wildfire} \Big) z_k^i + \beta_{FMC} 1 \left[d_t^i \neq J+1 \right] \cdot FMC_t^i + \epsilon_{jt}^i. \quad (12)$$

5 Results

5.1 Identifying Variation in Fire Hazard

A key assumption underlying our boundary discontinuity design approach is that unobserved variables correlated with wildfire hazard do not vary discontinuously at the disclosure requirement boundary. Because disclosure requirements are in part a function of hazard, hazard might increase discontinuously at that boundary. If so, unobserved variables correlated with wildfire hazard might do so as well. Although we cannot directly assess the validity of this assumption, Figure 3 provides some evidence of it by examining how wildfire hazard itself varies across the boundary; it provides a binned scatter plot of residuals of $\log(1 + WHP)$ within bins of distance within the boundary. $\text{Log}(1 + WHP)$ is residualized on boundary ID fixed effects to account for fixed differences in hazard across areas closest to different boundary segments. The figure indicates that although hazard increases with distance inside the boundary, as expected, it does not vary discontinuously at the boundary, suggesting that unobserved variables correlated with hazard may not either.

5.2 Aversion to Wildfire Hazard

Tables 4–5 summarize results from estimating Equation 1 based on Equations 11 and 12. The main parameter of interest is the effect of wildfire hazard disclosure on WTA for hazard ($\alpha_{k,DD}$). We focus on this because households are more likely to have full information about wildfire hazard for locations in disclosure-regulated areas. We estimate this parameter for individuals of three income types—low, medium, and high—based on terciles of the sample's income distribution.[20] Among low-income households, aversion to wildfire hazard in disclosed areas is $\alpha_{0,DD}$, which is the coefficient on $Disc_j \cdot \log(1 + WHP_j)$ (from Equation 11). For medium- and high-income households, aversion is given by $\alpha_{0,DD} + \alpha_{1,DD}$ and $\alpha_{0,DD} + \alpha_{2,DD}$, respectively; that is, preference parameter estimates for the medium $\alpha_{1,DD}$ and high $\alpha_{2,DD}$ groups should be interpreted as the additional effect of hazard disclosure on aversion to hazard relative to the low-income (or base) group. We present demand estimates from Stages 1 and 2 in a single table for the choice attributes for which individuals have heterogeneous preference (X_{2jt}).[21] All preference parameter estimates are converted into WTP/WTA values by dividing by the coefficient on moving costs (β_{FMC}), which is also shown in the tables.

Table 4 presents preference estimates from the base difference-in-differences specification. Households dislike moving costs, as shown by the negative parameter estimate for β_{FMC} (p<0.01). Household aversion to wildfire hazard is greater for homes in disclosure areas, where the minimum amount low-income households would be willing to accept in exchange for a 1 percent increase in WHP is \$258 per year; this translates to roughly 0.83 percent of the average flow cost of housing (based on an average home price of \$618,298 and a discount rate of 5 percent). The effect of disclosure on WTA values associated with wildfire hazard also increases with income, as WTA values among medium- and high-income households are 11 (\$29) and 20 (\$53) percent more per year than the base group, respectively. Without contrasting disclosure and nondisclosure areas, one might infer that households *like* higher wildfire risk areas. In nondisclosure areas, the estimated WTP for wildfire hazard is positive for all income groups, likely reflecting positive WTP for amenities correlated with hazard. In the base group, households are willing to pay \$62 for a 1 percent increase in wildfire hazard; medium- and high-income households are willing to pay \$25 and \$54 less, resulting in total WTP of \$37 and \$8, respectively. Consistent with previous work, however, disclosure changes the perceived riskiness of a location, possibly allowing individuals to make more-informed choices.[22]

[20] The cutoff between the first and second terciles is \$104 thousand, and the cutoff for the second and third terciles is \$182 thousand for the 1 km sample.

[21] We present the full set of estimates from Stage 2 for several specifications in Appendix Table A.1.

[22] Conditional on risk, individuals seem to be willing to pay for risk disclosure ($Disclosure$), and the amount they are willing to pay increases with income. However, it is also possible that the positive coefficients on regulation status captures a higher provision of public goods in areas with mandatory disclosure.

Estimates of $\alpha_{k,DD}$ provide our best estimates of preference for wildfire hazard because they capture the effect of disclosure, which both informs home buyers about hazards and makes them more salient, on aversion to hazard, while holding other factors—including correlated amenities—constant. Nevertheless, estimates of $\alpha_{k,DD}$ likely represent an underestimate of aversion to wildfire hazard because they do not include preferences over wildfire hazard common across properties inside and outside of disclosure zones. These preferences are captured by estimates of $\alpha_{k,r}$ (the coefficient on $\log(1+WHP_j)$). However, this coefficient potentially captures preferences for correlated amenities. Although our boundary-segment fixed effect and 1 km sample limitation are aimed to mitigate the impact of correlated amenities, the portion of preferences captured by $\alpha_{k,r}$ that are due to wildfire hazard as opposed to amenities is unknown, and we take a conservative approach.

Our findings also suggest additional aversion to areas with fire events, conditional on background risk. In Table 4, the average WTA for locations that have seen wildfire activity in the last five years is \$999–\$1,129 per year, depending on income, or about 3.2–3.7 percent of average housing prices. This may be driven by availability bias (Tversky & Kahneman, 1973) to recent events or because locations sustain undesirable residual effects from fires, such as burn scars (McCoy & Walsh, 2018). Although our identification strategy does not rely on wildfire events, a hedonic literature estimates their price impacts (Huggett, Murphy, & Holmes, 2008; Stetler, Venn, & Calkin, 2010; Loomis, 2004; McCoy & Walsh, 2018; Mueller, Loomis, & González-Cabán, 2009; Mueller & Loomis, 2008, 2014) and frequently finds immediate price impacts after a wildfire event of 10 percent or higher. For example, Loomis (2004) finds impacts of a wildfire in Colorado to be 15 percent, and Mueller, Loomis, & González-Cabán (2009) find price decreases of 10 percent due to wildfires in southern California, where additional fires yield even higher price reductions of 23 percent. Although the estimated price effects of events in our study are much lower, McCoy & Walsh (2018) find that price discounts from wildfires in Colorado disappear year after a fire. Aside from differences in geographic region/scope and time, the impact of wildfire events that we measure are within *five* years of the event, during which the price impacts, conditional on latent risk, could have diminished significantly.

To help gauge the size of our WTA estimates for fire hazard, we discretize WHP into bins of zero, low, and high and alter our specification to allow the effect of disclosure to differ across them.[23] This specification thus allows WTA to vary with hazard level. Table 5 presents the estimates. Focusing on WTA in disclosure areas, we find that it is only statistically significant in high-hazard areas: in areas with disclosure requirements, the lowest-income group's additional WTP to avoid living in a high-hazard area, relative to areas without disclosure requirements, is equivalent to 4.6 percent of the average flow cost of housing; the highest-income group's additional WTP is equivalent to 5.7 percent of housing costs. For low-hazard areas, the WTA monotonically decreases with income; that is, higher-income households care less about avoiding low-hazard areas than lower-income households do.

Last, the signs of the preference parameters for other characteristics are as expected across both the specifications in Tables 4 and 5: WTP values for lot size, school quality (Avg. percent meeting standard), and proximity to open areas (inverse of log(Distance to protected areas)) are positive and increase with income, which is as expected for normal goods. To the extent that the White population share captures higher amenities, the sign on Percent White non-Hispanic is also intuitive.

5.3 Heterogeneity by Fire Experience

We examine whether aversion to wildfire hazard varies according to a buyer's experience with fire at a previous residential location in Table 6. The majority of our sample (75 percent) within 1 km of disclosure area boundaries did not experience a wildfire event, and so estimated preferences for this group (Table 6, Columns 1–3) are similar to the overall estimates in Table 4. For the group *with* fire experience, we find that the effect of disclosure on WTP to avoid a marginal increase in wildfire risk is lower by about 4–5 percent (or $9–$15), conditional on income. For example, disclosure increases annual WTA for a 1-percent increase in risk in disclosure-regulated areas by $315 for buyers without fire experience in the highest-income tercile, whereas it increases annual WTA by $300 for high-income buyers with wildfire experience. The WTA for wildfire *events* by fire experience mirrors that for ex ante fire risk, where aversion to fire hazard is slightly lower for those with fire experience.

[23] Low-WHP areas have WHP greater than 0 and less than or equal to 401; high-WHP areas have scores greater than 401. These hazard cutoffs are applied based on Dillon, Menakis, & Fay (2014).

Specifically, for those without fire experience, WTA for living in a location that has experienced a recent fire is, on average, 3.5 percent of the average flow cost of housing across the three income groups. In contrast, WTA for those with fire experience ranges from 2.8 percent (for the highest-income group) to 3.5 percent (lowest-income tercile).

Fire experience seems to *reduce* WTA for wildfire hazard in areas where risks must be disclosed, although the impacts are small. The literature on the effects of disaster experience on location choice is thin, but the consensus is that demand to avoid a catastrophe is likely to increase with experience (González-Cabán et al., 2013; Cohen, Etner, & Jeleva, 2008; Cameron & Englin, 1997). González-Cabán et al. (2013) conduct a choice experiment and finds that aversion to wildfire loss is driven by those with personal experience with wildfire impacts, defined by whether the individual or a family member suffered health damages from smoke or travel was affected.[24] Hyde (2023) documents that homeowners who experience flooding tend to shift purchases away from coastal areas in Florida in interstate relocation decisions.

Personal experience with disasters may also shift decisions that affect location choice. Experience with catastrophic risk is likely to increase individual demand for insurance. Cohen, Etner, & Jeleva (2008) predicts this to be the case assuming a rank-dependent expected utility model. Gallagher (2014) empirically documents spikes in flood insurance takeup in communities that recently experienced flooding. Gallagher & Hartley (2017) show that flooding causes individuals to pay off their home mortgage using insurance payments rather than rebuild, which suggests that personal experience pushed individuals *away* from their flood-prone location.

Our results on the effect of experience depart from previous work. Several mechanisms may explain this. If individuals learn in a Bayesian fashion, then lower-than-expected realized damages from a previous fire experience may cause them to revise their risk beliefs downward, lowering overall posterior beliefs regarding risks and thus aversion to risk. There may also be a "learning by doing" effect, where individuals learn how to better navigate insurance and

[24] This study differs from ours. First, "personal experience" in González-Cabán et al. (2013) is specified to yield damage (in terms of health or travel), whereas that may not be the case in our setting. Second, revealed preference differs from stated preference methodology. Cameron & Englin (1997) find that experience generally increases WTA in contingent valuation settings.

mitigation (which can reduce insurance costs) with wildfire experience. Finally, households that have experienced a wildfire nearby may have underlying preferences to live in locations where such an event is likely. Although we cannot disentangle the specific mechanism, what remains clear is that experience (as a source of information on risks) in not a replacement for public risk disclosure. When comparing preference estimates for risk by disclosure-regulation status, aversion to ex ante risk in nondisclosure areas for those with fire experience remains positive overall and similar to those without fire experience. Taken together, individual experience with fire does not seem to be a major deterrent to living in wildfire hazard areas and, if anything, renders individuals less averse. This finding is based on individual location choices in our sample period of 2009–2018. Since then, regulatory and market incentives have changed. For example, wildfire insurance is now more difficult to obtain, with major insurance companies scaling back coverage in several disaster-prone areas. If the effects of fire experience that we measure are driven by the structure of insurance, then individual experience may have a very different impact on location choices going forward. On the other hand, this may not be the case if individuals tend to overreact to low-probability events and fire experience simply corrects their risk perceptions. Understanding the drivers of fire experience would be a valuable avenue of future research for predicting wildfire damages as these risks continue to grow and affect a larger share of the population.

5.4 Robustness

We examine the robustness of our findings by either augmenting the estimation model or modifying the estimate sample.

Full Sample of Choices Although our main analysis focuses on areas within 1 km of disclosure area boundaries for identification, doing so may arbitrarily limit the choice set individuals face (to areas near boundaries). To test how this affects our estimates, we re-estimate the model and include *all* areas within California but additionally distinguish choices near versus far from boundaries. To control for correlates of wildfire risk, we then alter the specification to allow for a full set of interactions between disclosure-regulation status, log of WHP, and a 1 km boundary indicator. As in our main estimates, we include boundary segment fixed effects (BD_j, in Equation 11). However, in this specification, the boundary segment fixed effects are less powerful as controls, as they capture similarities between properties that are nearest to a given segment but potentially farther than 1 km away. Therefore, in addition to these, we now include a set of proximity to boundary by boundary segment fixed effects; as in our main estimates, these fixed effects capture time-invariant similarities in the value of properties *within 1 km* of the same segment.[25] The coefficient on the triple interaction returns the difference in the effect of disclosure on preferences over hazard between areas that are near versus far away from the disclosure area boundary. The measured effect of disclosure on preferences over hazard within areas far from the boundary ($Disclosure \times \log(1 + WHP)$) should capture preferences for amenities that differ between disclosure and nondisclosure areas and are correlated with hazard. After parceling out this effect, the effect measured in the triple interaction should be comparable to the parameter of interest in our main model. Table 7 presents the results. The estimated values of WTA for fire hazard (ranging from \$228 to \$262) are generally similar to our main estimates.

Subsample Analysis A concern is that disclosure regulation occurs in areas of higher wildfire hazard. Our boundary discontinuity strategy should alleviate this concern to some extent, given that WHP transitions continuously across disclosure area boundaries (as shown in Figure 3). Moreover, we also control for differences across FHSZs in all specifications through hazard zone fixed effects. Still, we can test the effect of this concern by additionally limiting our analysis to just one hazard zone, which still has variation in regulation and wildfire risk. Table 8 presents results that limit the estimation sample to areas rated as high hazard: high FHSZs in SRAs and LRAs classified by CAL FIRE as high hazard within its nine-class hazard map. High FHSZ SRA properties face the disclosure requirement, whereas properties in the LRAs do not. Households in the lowest-income tercile are willing to pay 0.98 percent of home values (or \$305) to avoid a 1 percent increase (compared to \$258 in

[25] Although our main estimates include N_{BD} boundary fixed effects, where N_{BD} is the number of boundary segments that properties in our sample are nearest to, this triple difference specification includes $N_{BD} \times 2$ boundary fixed effects: a set of main effects and a set of effects that apply only to properties within 1 km of their closest boundary segment. The latter set is collinear with the base group's *within1km* parameter; therefore, an estimate for this parameter does not appear in Table 7.

the full sample). Interestingly, when focusing on high-hazard areas, we find that preferences over wildfire hazard are negative, even outside of disclosure areas. There, the minimum value a household would be willing to accept for a 1 percent increase in hazard is \$186–\$225 compared to -\$62–\$8 in our main estimates that include all hazard areas. This is a somewhat encouraging result in that individuals are less likely to rely on mandatory disclosure in areas where hazard risks are high.

Similarly, another concern is that the impacts of disclosure that we estimate may be picking up differences across responsibility areas, which also roughly correspond with a location's incorporation status. To assess this concern, we re-estimate our main specification using properties in LRA only, where most locations are incorporated. Disclosure requirements vary across properties within and outside very-high FHSZ boundaries (although, as we show in Figure 1, hazard varies substantially within and across hazard zones). Table 9 presents these results. The effect of disclosure on WTP to avoid fire risk in disclosure-required areas are similar to the main estimates.

Last, Chapter 7A building codes were strengthened in 2008 to improve the fire resistance and management of high-risk properties. Updated codes applied to homes in disclosure-regulated areas that were built in 2008 or later. To remove any confounding of building codes with disclosure, we remove all properties built in 2008 or later and present estimates in Table 10. The estimated effect of disclosure on WTP to avoid risk in this sample (-\$195 to -\$261 per year) is slightly lower in magnitude than the main estimates (-\$258 to -\$311 per year). With buildings being less fire resistant, one might expect the magnitude of the WTP to avoid fire risk to increase rather than decrease. Our results largely reflect development, after 2007, into higher fire risk areas: the average WHP for properties built in 2008 or later is 556, compared to 191 for properties built before 2008. The increased risk seems to more than offset regulated building code improvements in fire resistance.

6 Discussion

Over the past several decades, what it means to live in a US wildfire hazard area has changed dramatically. Risks have escalated. Homeowners face challenges getting insurance, as evidenced by increasing takeup of FAIR plans and, in California, the withdrawal of major insurance companies. Households in high-hazard areas are more likely to be high income. However, this may in part be due to preferences among higher-income households for wildland amenities that are spatially correlated with hazard. Increasing hazard and the changing landscape of regulatory and market incentives may influence who chooses to live in high-hazard areas. But to understand this, it is necessary to understand preferences about hazard itself (separate from correlated amenities) and how those vary across income groups.

We use spatial discontinuity in California's wildfire hazard disclosure laws to understand aversion to hazard and how it varies with income, fire experience, and information. Comparing preferences for nearby homes for which disclosure laws make fire hazard more and less salient allows us to specifically identify effects of information on WTA, while accounting for fixed neighborhood characteristics that are common to neighboring homes where disclosure is and is not required. We account for price endogeneity using a strategy based on moving costs, which also allows us to investigate the relationship between experience with wildfires and aversion to wildfire hazard.

Our main results indicate that, relative to homes in areas where disclosure is not required, households in the lowest-income tercile are willing to pay $258 less per year for each percent increase in wildfire hazard for homes where disclosure is required. This value equals 0.8 percent of the estimated cost of housing in our study region. Similar to findings in Donovan, Champ, & Butry (2007), this aversion to hazard exists only in areas where disclosure is required; elsewhere, estimates indicate that households have a preference for or are neutral toward hazard—possibly due to preferences for correlated amenities. Moreover, aversion to hazard is increasing in income. The effect of disclosure on WTA values for a 1 percent change in wildfire hazard is 12 percent greater in the highest- versus lowest-income tercile. These results suggest potential concerns around distributional equity across income groups as hazard increases, with lower-income households replacing higher-income households over time in high-hazard locations.

To our knowledge, these are the first estimates of sorting over wildfire hazard; however, they mirror results in other contexts. Bakkensen & Ma (2020) found that aversion to living in high-risk flood zones was 14 percent higher among the highest- than lowest-income quintile of households in their Florida sample; aversion was 32 percent lower among Black as compared to White households. In a model of interregional sorting, Fan, Fisher-Vanden, & Klaiber (2018) found that the minimum value an individual would be willing to accept for an additional extreme heat day was 50 percent higher among individuals with a college degree than those without. It is somewhat difficult to make direct comparisons between WTA values associated with a 1 percent increase in wildfire hazard, a high-risk flood zone, and an additional day of extreme heat. However, our estimates appear similar in magnitude based on results in Table 5. We find that disclosure increases aversion toward high-WHP areas by about 24 percent in the highest- versus lowest-income tercile. For the lowest-income group, we find that disclosure increases WTA values associated with high WHP areas by about \$1,400 per year—about 4.6 percent of the average flow cost of housing in our sample; Bakkensen & Ma (2020) found that within their base group, aversion to high-risk flood zones was equivalent to 6.3 percent of average annual housing costs.

Our results also point to the potentially important role information provision may have in shaping responses to increasing wildfire hazard. We find that households *prefer* homes with higher levels of wildfire hazard when disclosure is not required, likely owing to imperfect information and preferences for wildland amenities that are correlated with hazard. However, WTA values associated with hazard become negative when an NHD focuses households' attention on potential negative side effects of those amenities. Only California and Oregon have laws requiring disclosure in home sales; however, these results suggest that disclosure laws could be an important piece of the puzzle in discouraging relocation into potentially dangerous areas. Unfortunately, our results also suggest that disclosure laws may have unequal effects: disclosure reduces higher-income households' aversion toward high-hazard areas more strongly than it does for low-income households. This suggests that although disclosure laws may be well justified from the standpoint of procedural justice—it is important for households to understand what they are buying—they may have consequences for distributional justice, with lower-income households replacing high-income households in high-hazard areas.

Last, we provide what are, to our knowledge, the first estimates of the effects of hazard experience on locational sorting, enabled by tracking individual location decisions over time.

Experience matters in individual choice across a broad range of consumer markets. For the residential market, the potential magnitude of wildfire impacts suggests that any such experience would be a significant deterrent to living in the WUI. Instead, our results indicate that (1) experience does not substantially alter WTP in areas where disclosure is not required, and (2) those with fire experience have less aversion to fire hazard than those without in disclosure areas. First and foremost, our result implies that individual experience with wildfire is not a replacement for disclosure regulation, which enables more-informed decision making. That is, one cannot rely on an increasing population in the WUI (or growing knowledge of the impacts of risk) to correct information failures regarding the properties at risk of fire damage. Our second result suggests that additional work to understand *how* fire experience facilitates decisions regarding risk exposure is required. Two potential explanations include a correction of risk perceptions from an overreaction to potential wildfire hazard or learning-by-doing that facilitates damage mitigation, through actual risk mitigation or better navigation of the insurance claims process. If experience reduces the costs of damage mitigation, then the effects of experience with wildfire that we measure may no longer materialize going forward with changes in the insurance market. How fire experience affects the choice to live in the WUI will depend on the explanation and is crucial for assessing future wildfire damages as population and development in the WUI continue to grow.

References

Ackerberg, D. A. (2003). Advertising, learning, and consumer choice in experience good markets: an empirical examination. International Economic Review, 44, 1007–1040.

Ashworth, J., Hotz, V. J., Maurel, A., & Ransom, T. (2021). Changes across cohorts in wage returns to schooling and early work experiences. Journal of labor economics, 39, 931–964.

Bakkensen, L. A. & Ma, L. (2020). Sorting over flood risk and implications for policy reform. Journal of Environmental Economics and Management, 104, 102362.

Bayer, P., Keohane, N., & Timmins, C. (2009). Migration and hedonic valuation: The case of air quality. Journal of Environmental Economics and Management, 58, 1–14.

Bayer, P., McMillan, R., Murphy, A., & Timmins, C. (2016). A dynamic model of demand for houses and neighborhoods. Econometrica, 84, 893–942.

Baylis, P. & Boomhower, J. (2023). The economic incidence of wildfire suppression in the United States. American Economic Journal: Applied Economics, 15, 442–473.

Baylis, P. W. & Boomhower, J. (2022). Mandated vs. voluntary adaptation to natural disasters: The case of US wildfires. Technical report, National Bureau of Economic Research.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. Econometrica: Journal of the Econometric Society, (841–890).

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. The RAND Journal of Economics, 25, 242–262.

Billings, S. (2019). Technical summary - merging home mortgage disclosure data to property records from Zillow.

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

Bishop, K. C. & Murphy, A. D. (2011). Estimating the willingness to pay to avoid violent crime: A dynamic approach. American Economic Review, 101, 625–629.

Bishop, K. C. & Murphy, A. D. (2019). Valuing time-varying attributes using the hedonic model: when is a dynamic approach necessary? Review of Economics and Statistics, 101, 134–145.

California Legislature (1997). Assembly bill no. 1195, 1997-98 regular session. Available: `http://www.leginfo.ca.gov/pub/97-98/bill/asm/ab_1151-1200/ab_1195_bill_19980521_amended_sen.pdf`.

Cameron, T. A. & Englin, J. (1997). Respondent experience and contingent valuation of environmental goods. Journal of Environmental Economics and Management, 33, 296–313.

Chernew, M., Gowrisankaran, G., & Scanlon, D. P. (2008). Learning and the value of information: Evidence from health plan report cards. Journal of Econometrics, 144, 156–174.

Cignarale, T., Laucher, J., Allen, K., & Landsman-Smith, L. (2017). The availability and affordability of coverage for wildfire loss in residential property insurance in the wildland-urban interface and other high-risk areas of California: CDI summary and proposed solutions. California Department of Insurance's Availability and Affordability Residential Property Insurance Task Force.

Cohen, M., Etner, J., & Jeleva, M. (2008). Dynamic decision making when risk perception depends on past experience. Theory and Decision, 64, 173–192.

Dillon, G. K., Menakis, J., & Fay, F. (2014). Wildland fire potential: A tool for assessing wildfire risk and fuels management needs. In Proceedings of the large wildland fires conference, (60–76).

Donovan, G. H., Champ, P. A., & Butry, D. T. (2007). Wildfire risk and housing prices: a case study from colorado springs. 83, 217–233. ISBN: 0023-7639 Publisher: University of Wisconsin Press.

Eidenshink, J., Schwind, B., Brewer, K., Zhu, Z.-L., Quayle, B., & Howard, S. (2007). A project for monitoring trends in burn severity. Fire ecology, 3, 3–21.

Fan, Q. & Davlasheridze, M. (2015). Flood risk, flood mitigation, and location choice: evaluating the national flood insurance program's community rating system. Risk Analysis, 36, 1125–1147.

Fan, Q., Fisher-Vanden, K., & Klaiber, H. A. (2018). Climate change, migration, and regional economic impacts in the United States. Journal of the Association of Environmental and Resource Economists, 5, 643–671.

Fan, Q., Klaiber, H. A., & Fisher-Vanden, K. (2016). Does extreme weather drive interregional brain drain in the US? Evidence from a sorting model. Land Economics, 92, 363–388.

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

Gallagher, J. & Hartley, D. (2017). Household finance after a natural disaster: The case of hurricane katrina. American Economic Journal: Economic Policy, 9, 199–228.

Garnache, C. (2023). Does the salience of climate-related risk affect asset prices? Available at SSRN: https://ssrn.com/abstract=3398404 or http://dx.doi.org/10.2139/ssrn.3398404.

Gibson, M. & Mullins, J. T. (2020). Climate risk and beliefs in New York floodplains. Journal of the Association of Environmental and Resource Economists, 7, 1069–1111.

González-Cabán, A., Holmes, T. P., Loomis, J. B., & Sánchez, J. J. (2013). Does personal experience affect choice-based preferences for wildfire protection. In International Symposium on Fire Economics, Planning, and Policy: Climate Change and Wildfires, (111).

Greenstone, M. (2017). The continuing impact of sherwin rosen's "hedonic prices and implicit markets: product differentiation in pure competition". Journal of Political Economy, 125, 1891–1902.

Hallstrom, D. & Smith, V. (2005). Market responses to hurricanes. Journal of Environmental Economics and Management, 50, 541–561.

Higuera, P. E., Cook, M. C., Balch, J. K., Stavros, E. N., Mahood, A. L., & St. Denis, L. A. (2023). Shifting social-ecological fire regimes explain increasing structure loss from western wildfires. PNAS nexus, 2, pgad005.

Hino, M. & Burke, M. (2021). The effect of information about climate risk on property values. Proceedings of the National Academy of Sciences, 118, e2003374118.

Huggett, R. J., Murphy, A., & Holmes, P. (2008). Forest disturbance impacts on residential property values. The Economics of Forest Disturbances: Wildfires, Storms, and Invasive Species., (209–228).

Hyde, T. (2023). How do homebuyers adapt after experiencing a natural disaster? evidence from the florida real estate market. Working Paper.

Insurance Information Institute (2023). Facts + statistics: U.S. catastrophes. Available at: `https://www.iii.org/fact-statistic/facts-statistics-us-catastrophes`.

Issler, P., Stanton, R., Vergara-Alert, C., & Wallace, N. (2020). Mortgage markets with climate-change risk: Evidence from wildfires in California. Available at SSRN 3511843.

Kousky, C. (2010). Learning from extreme events: Risk perceptions after the flood. 86, 395–422. ISBN: 0023-7639 Publisher: University of Wisconsin Press.

Kramer, H. A., Mockrin, M. H., Alexandre, P. M., & Radeloff, V. C. (2019). High wildfire damage in interface communities in California. International Journal of Wildland Fire, 28, 641–650.

Liao, Y., Walls, M., Wibbenmeyer, M., & Pesek, S. (2022). Insurance availability and affordability under increasing wildfire risk in California. Resources for the Future: Issue Brief, (22–09).

Loomis, J. (2004). Do nearby forest fires cause a reduction in residential property values? Journal of Forest Economics, 10, 149–157.

Ma, L. (2019). Learning in a hedonic framework: Valuing brownfield remediation. International Economic Review, 60, 1355–1387.

Ma, L., Walls, M., Wibbenmeyer, M., & Lennon, C. (2024). Risk disclosure and home prices: Evidence from California wildfire hazard zones. Land Economics, 100, 6–21.

Mann, M. L., Berck, P., Moritz, M. A., Batllori, E., Baldwin, J. G., Gately, C. K., & Cameron, D. R. (2014). Modeling residential development in California from 2000 to 2050: Integrating wildfire risk, wildland and agricultural encroachment. Land use policy, 41, 438–452.

McCoy, S. J. & Walsh, R. P. (2018). Wildfire risk, salience & housing demand. 91, 203–228. ISBN: 0095-0696 Publisher: Elsevier.

McFadden, D. (1978). Modelling the choice of residential location. Institute of Transportation Studies, University of California.

Miller, R. K., Field, C. B., & Mach, K. J. (2020). Factors influencing adoption and rejection of fire hazard severity zone maps in California. International Journal of Disaster Risk Reduction, 50, 101686.

Mueller, J., Loomis, J., & González-Cabán, A. (2009). Do repeated wildfires change homebuyers' demand for homes in high-risk areas? a hedonic analysis of the short and long-term effects of repeated wildfires on house prices in southern California. The Journal of Real Estate Finance and Economics, 38, 155–172.

Mueller, J. M. & Loomis, J. B. (2008). Spatial dependence in hedonic property models: Do different corrections for spatial dependence result in economically significant differences in estimated implicit prices? Journal of Agricultural and Resource Economics, (212–231).

Mueller, J. M. & Loomis, J. B. (2014). Does the estimated impact of wildfires vary with the housing price distribution? a quantile regression approach. Land Use Policy, 41, 121–127.

Nolte, C., Boyle, K. J., Chaudhry, A., Clapp, C., Guignet, D., Hennighausen, H., Kushner, I., Liao, Y., Mamun, S., Pollack, A., Richardson, J., Sundquist, S., Swedberg, K., & Uhl, J. H. (2023). Studying the impacts of environmental amenities and hazards with nationwide property data: Best data practices for interpretable and reproducible analyses.

Radeloff, V. C., Helmers, D. P., Kramer, H. A., Mockrin, M. H., Alexandre, P. M., Bar-Massada, A., Butsic, V., Hawbaker, T. J., Martinuzzi, S., Syphard, A. D., et al. (2018). Rapid growth of the US wildland-urban interface raises wildfire risk. Proceedings of the National Academy of Sciences, 115, 3314–3319.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. Journal of political economy, 82, 34–55.

Starrs, C. F., Butsic, V., Stephens, C., & Stewart, W. (2018). The impact of land ownership, firefighting, and reserve status on fire probability in California. Environmental research letters, 13, 034025.

Stetler, K. M., Venn, T. J., & Calkin, D. E. (2010). The effects of wildfire and environmental amenities on property values in northwest Montana, USA. Ecological Economics, 69, 2233–2243.

Swiss Re Institute (2023). In 5 charts: continued high losses from natural catastrophes in 2022. `https://www.swissre.com/institute/research/sigma-research/sigma-2023-01/5-charts-losses-natural-catastrophes.html`. Accessed: January 16, 2024.

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

US Social Security Administration (1969). Russell Soundex Coding and Filing System: A Programmed Text. U.S. Government Printing Office.

van der Loo, M. (2014). The stringdist package for approximate string matching. The R Journal, 6, 111–122.

Wibbenmeyer, M. & Robertson, M. (2022). The distributional incidence of wildfire hazard in the western United States. Environmental Research Letters, 17, 064031.

Zivin, J. G., Liao, Y., & Panassie, Y. (2023). How hurricanes sweep up housing markets: Evidence from Florida. Journal of Environmental Economics and Management, 118, 102770.

Table 1: Demographic Effects of Limiting the Sample to Transactions Matched with HMDA Loan Records and Origin Home Sales Transactions

Variable	Census (homeowners)	HMDA	HMDA-ZTRAX	HMDA-ZTRAX-Origin Home
Mean Percent White	79.87	76.19	78.38	78.77
Mean Percent Black	4.97	3.97	2.32	2.17
Mean Percent Asian	14.05	18.79	17.60	17.38
Mean Percent American Indian	0.85	1.04	1.04	1.05
Mean Percent Native Hawaiian Pacific Islander	0.26	0.93	0.66	0.63
Mean Percent Hispanic (Ethnicity)	21.41	21.58	26.38	27.02
Median Household Income (Thousands of Dollars)	94.15	105	110	111

Note: Race and ethnicity statistics in Column 1 describe the population of California homeowners based on data from the 2010 Decennial Census. We exclude other and multiracial categories; therefore, race categories sum to 100 percent. Median income data in Column 1 are based on the 2013 American Community Survey 5-year estimate of median household income for households with a mortgage. Column 2 summarizes demographic characteristics in California HMDA data, after limiting the data set to completed loans used for new property purchases, as described in Appendix B.2. Column 3 provides summary statistics for the set of ZTRAX transactions successfully matched to HMDA records. The sample in Column 4 is further limited to those for which an origin home sale was successfully identified. Depending on the year of collection, the HMDA LAR combines Pacific Islander and Asian into a single category—Asian and Pacific Islander, coded as 2 on the loan register—or two categories—Asian, coded as 2, and Native Hawaiian Pacific Islander, coded as 4. We report all 2 codes as "Asian" and all 4 codes as "Native Hawaiian Pacific Islander."

Table 2: Summary Statistics (mean and standard error) for Property Characteristics Within the Final HMDA and Origin Home Matched Transaction Data Set

Variable	Full Sample	Regulated	Unregulated
Sales price amount	618,297.98	778,535.56	585,990.06
	1,090.96	3,058.31	1,138.48
Bedrooms	3.49	3.55	3.47
	0.002	0.006	0.002
Lot size (acres)	0.90	1.31	0.82
	0.047	0.027	0.056
Building area (square feet)	2,084.77	2,525.29	1,995.95
	2.87	6.04	3.18
Age of home	37.87	30.58	39.34
	0.06	0.13	0.07
Wildfire Hazard Potential	223.39	943.73	78.15
	3.10	16.44	1.42
Percent White non-Hispanic	50.30	68.56	46.62
	0.06	0.11	0.07
Avg. percent meeting standard	48.38	51.75	47.70
	0.03	0.09	0.04
Previous fire within 5 years	0.02	0.11	0.01
	0.0004	0.002	0.0002
Distance to protected areas (meters)	15,428.3	11,052.35	16,310.6
	29.79	50.22	33.87
Incorporated area	0.81	0.52	0.87
	0.001	0.003	0.001
Number of obs.	177,247	29,740	147,507

Note: Sales price amount, number of bedrooms, lot size, and age of home come from ZTRAX. Wildfire Hazard Potential at the location of each home is based on Dillon, Menakis, & Fay (2014). Percent White non-Hispanic is based on 2010 tract-level decennial Census data. Average percent meeting standards is the average of the percentage of students in the property's school district meeting math and language arts standards and from the California Department of Education. Distance to protected areas is the distance to the nearest area classified under GAP Status 1 or 2 within the USGS Protected Area Database. Incorporated area is based on data from CAL FIRE.

Table 3: Property Summary Statistics (mean and standard error), by Buyer Characteristics

Variable	Income Q1	Income Q2	Income Q3	No Fire Exp.	Fire Exp.
Sales price amount	349,868.07	606,850.44	1,161,945.75	618,712.81	616,473.06
	589.48	1,073.73	2,810.45	1,224.42	2,384.51
Bedrooms	3.26	3.55	3.84	3.48	3.49
	0.003	0.004	0.005	0.002	0.005
Building area (square feet)	1,701.9	2,139.08	2,767.75	2,072.62	2,138.05
	2.14	7.14	5.22	3.32	5.12
Age of home	39.14	36.98	36.54	38.14	36.68
	0.09	0.10	0.13	0.07	0.14
Percent White non-Hispanic	41.11	54.21	63.27	50.01	51.57
	0.09	0.10	0.10	0.07	0.14
Wildfire Hazard Potential	214.24	246.76	210.81	209.76	283.14
	4.53	5.97	5.93	3.32	8.14
Disclosure	0.11	0.18	0.26	0.15	0.22
	0.001	0.002	0.002	0.001	0.002
Previous fire within 5 years	0.01	0.03	0.04	0.02	0.04
	0.0004	0.0007	0.0009	0.0004	0.001
Number of obs.	81,596	54,221	41,430	144,339	32,908

Note: Variables are as defined in Table 2. Income columns Q1, Q2, and Q3 summarize the property characteristics of homes bought by households in the first, second, and third terciles of income, respectively. Fire experience columns summarize the property characteristics of homes bought by households that did not and did, respectively, experience a nearby fire (within 5 km) at their previous home.

Table 4: Heterogeneous Preferences for Location Attributes

	Base	Q2	Q3
Disclosure	1,072.51	149.58	265.97
	487.69	13.16	13.79
log(1+WHP)	61.99	-25.74	-53.91
	59.42	3.13	3.62
Disclosure × log(1+WHP)	-257.56	-29.45	-53.37
	78.39	4.14	4.62
log(Lot size)	3,561.62	31.59	88.64
	175.82	6.94	7.35
Avg. percent meeting standard	12,645.10	1,266.52	2,144.13
	2,213.47	34.03	35.99
Percent White non-Hispanic	8,463.63	324.17	862.06
	1,354.52	21.77	26.24
log(Distance to protected areas)	-1,897.20	-4.60	-60.68
	446.83	5.14	5.30
Built after 2008	2,169.53	203.59	507.16
	496.13	22.43	23.43
Southern California	-	2.42	-129.04
	-	11.50	12.03
Previous fire within 5 years	-1,129.13	119.41	130.16
	772.94	25.68	27.07
MC (β_{FMC})	-2.8014	0.0033	
Individuals ($\sum_t N_t$)	46,406		
Choices ($\sum_t J_t$)	12,919		

Note: The table presents estimates of preferences for location attributes for which buyers have heterogeneous preferences (by income). The specification follows Equation 11, where estimates from Columns 2 and on are from Stage 1 and estimates in Column 1 (base group) are from Stage 2. Second-stage additional controls (not shown) include the fixed effects by year, county-by-year, ZIP, boundary segment, and fire hazard class (moderate, high, very high). All individuals are assumed to have homogeneous preferences for the following house characteristics (not shown): bedrooms, age, log of square feet, and incorporation status. The sample includes locations within 1 km of disclosure boundaries across California. Base group preferences for *Southern California* are absorbed into county-by-year fixed effects within Stage 2 estimates and so not reported here.

Table 5: Heterogeneous Preferences for Location Attributes

	Base	Q2	Q3
Disclosure	810.29	33.43	64.12
	502.95	19.21	20.24
Low WHP (0 < WHP < 401)	85.87	-24.58	-53.55
	247.19	11.17	12.22
High WHP (WHP $\geq$ 401)	438.05	-118.74	-273.05
	509.12	29.46	35.16
Disclosure $\times$ Low WHP	-547.81	58.17	95.28
	394.34	22.16	23.27
Disclosure $\times$ High WHP	-1,413.18	-145.44	-343.56
	655.13	37.43	43.92
log(Lot size)	3,528.02	13.53	57.79
	175.27	6.68	7.10
Avg. percent meeting standard	1,2852.42	1260.56	2,122.59
	2215.26	34.08	35.87
Percent White non-Hispanic	8,578.20	313.19	860.00
	1,355.97	21.82	26.24
log(Distance to protected areas)	-1,872.41	-6.20	-68.03
	447.19	5.15	5.30
Built after 2008	2,182.47	144.68	400.53
	494.71	21.63	22.50
Southern California	-	9.92	-107.00
	-	11.46	11.93
Previous fire within 5 years	-1,200.78	105.39	90.18
	773.57	25.53	27.06
MC (β_{FMC})	-2.8052	0.0033	
Individuals ($\sum_t N_t$)	46,406		
Choices ($\sum_t J_t$)	12,919		

Note: This table re-estimates the model from Table 4, using the same sample, but discretizes wildfire hazard potential into zero, low, and high bins.

Table 6: Heterogeneous Preferences for Location Attributes with Fire Experience

	No Fire Experience			Fire Experience		
	Base	Q2	Q3	Q1	Q2	Q3
Disclosure	1,072.97	152.20	265.62	-0.07	145.39	275.84
	487.69	15.17	15.67	22.17	20.87	21.03
log(1+WHP)	63.76	-28.20	-56.26	-6.67	-24.84	-53.74
	59.42	3.57	4.04	4.82	5.25	6.18
Disclosure × log(1+WHP)	-260.37	-29.41	-54.51	11.18	-19.44	-39.75
	78.39	4.77	5.22	6.44	6.78	7.54
log(Lot size)	3,558.92	19.66	69.70	8.72	79.71	158.32
	175.82	8.00	8.34	11.18	11.09	11.08
Avg. percent meeting standard	12,629.90	1,234.77	2,143.86	68.49	1,475.42	2,280.84
	2,213.47	38.70	40.27	59.09	54.66	56.44
Percent White non-Hispanic	8,444.72	345.34	899.21	68.51	337.63	841.79
	1,354.52	24.93	29.68	34.20	36.89	44.88
log(Distance to protected areas)	-1,899.43	-5.20	-61.14	5.12	3.89	-52.81
	446.83	5.84	5.92	8.51	8.55	8.34
Built after 2008	2,168.02	221.00	535.62	2.35	156.89	430.36
	496.13	25.66	26.24	36.82	35.70	35.68
Southern California	-	-28.82	-164.25	247.10	414.52	325.56
	-	12.62	13.10	21.01	24.19	23.75
Previous fire within 5 years	-1,150.74	96.74	86.72	75.99	226.30	269.82
	772.94	31.12	32.38	42.41	37.06	37.52
MC (β_{FMC})	-2.7887	0.0033				
Individuals ($\sum_t N_t$)	46,406					
Choices ($\sum_t J_t$)	12,919					

Note: This table re-estimates the model from Table 4, using the same sample, but allows individual preferences to vary by fire experience.

Table 7: Robustness—Triple Difference

	Base	Q2	Q3
Disclosure	732.51	109.04	320.39
	407.78	16.17	16.57
log(1+WHP)	-34.61	-28.42	-64.73
	50.55	2.12	2.65
Within 1km of boundary	-	36.79	79.16
	-	6.79	7.73
Disclosure × log(1+WHP)	-58.00	2.60	-19.70
	74.30	3.49	3.94
Disclosure × Within 1km	391.81	24.93	-81.48
	410.68	19.89	20.61
log(1 + WHP) × Within 1km	132.32	12.25	24.54
	71.68	3.21	3.84
Disclosure × log(1+WHP) × Within 1km	-228.33	-34.19	-30.53
	101.18	5.01	5.66
log(Lot size)	3,590.04	7.02	51.60
	101.23	3.65	4.09
Avg. percent meeting standard	9,143.99	1,367.19	2,363.53
	1,309.57	16.19	18.25
Percent White non-Hispanic	6,551.35	439.18	824.76
	737.87	9.51	12.12
log(Distance to protected areas)	-1,188.18	-27.42	-69.06
	253.79	2.43	2.77
Built after 2008	2,477.36	94.72	332.66
	304.49	11.79	13.02
Southern California	-	-28.29	-144.79
	-	4.25	5.04
Previous fire within 5 years	-267.50	140.29	161.34
	474.05	16.86	18.16
MC (β_{FMC})	-3.1714	0.0020	
Individuals ($\sum_t N_t$)	177,250		
Choices ($\sum_t J_t$)	30,579		

Note: This table estimates preferences to avoid wildfires using the full sample of data, including choices made over locations beyond 1 km from the regulatory boundary. The model is modified to include a full set of interactions between regulated areas, log(1+WHP), and location within 1 km of a boundary.

Table 8: Robustness—High FHSZ only

	Base	Q2	Q3
Disclosure	374.53	-3.87	-55.91
	1,068.86	60.98	67.78
log(1+WHP)	-186.07	-14.98	-39.12
	117.10	6.39	7.48
Disclosure × log(1+WHP)	-305.19	-28.36	-63.24
	203.34	12.85	14.67
log(Lot size)	4,361.25	51.34	212.62
	442.65	17.73	19.38
Avg. percent meeting standard	6,839.65	1,752.77	3,258.76
	5,088.15	92.72	103.31
Percent White non-Hispanic	-693.19	244.87	489.47
	3,407.84	65.15	77.87
log(Distance to protected areas)	-2,380.02	-2.32	-69.67
	1,122.73	13.03	13.96
Built after 2008	3,460.86	186.79	440.75
	1,176.84	51.57	56.24
Southern California	-	-122.34	-393.46
	-	27.30	29.44
Previous fire within 5 years	-722.36	193.58	372.66
	2,287.28	71.49	72.97
MC (β_{FMC})	-2.4691	0.0076	
Individuals ($\sum_t N_t$)	8,274		
Choices ($\sum_t J_t$)	2,853		

Table 9: Robustness—Preference Estimates Using Houses in LRA Only

	Base	Q2	Q3
Disclosure	579.72	214.18	396.78
	389.61	19.99	21.15
log(1+WHP)	5.83	-13.45	-39.77
	76.16	4.77	5.43
Disclosure $\times$ log(1+WHP)	-270.67	-23.25	-38.60
	105.93	6.85	7.36
log(Lot size)	4,498.40	34.61	208.19
	284.14	13.54	13.70
Avg. percent meeting standard	14,640.18	1,512.86	2,357.52
	3,105.13	54.93	56.98
Percent White non-Hispanic	8,225.94	290.50	845.27
	2,081.40	36.96	42.52
log(Distance to protected areas)	-918.62	-6.03	-57.83
	712.67	8.58	8.71
Built after 2008	2,761.14	219.41	519.69
	712.42	36.20	37.50
Southern California	-	-63.33	-275.19
	-	19.51	19.91
Previous fire within 5 years	-1,786.46	136.93	137.17
	1,008.25	38.61	39.76
MC (β_{FMC})	-2.4695	0.0043	
Individuals ($\sum_t N_t$)	23,979		
Choices ($\sum_t J_t$)	7,107		

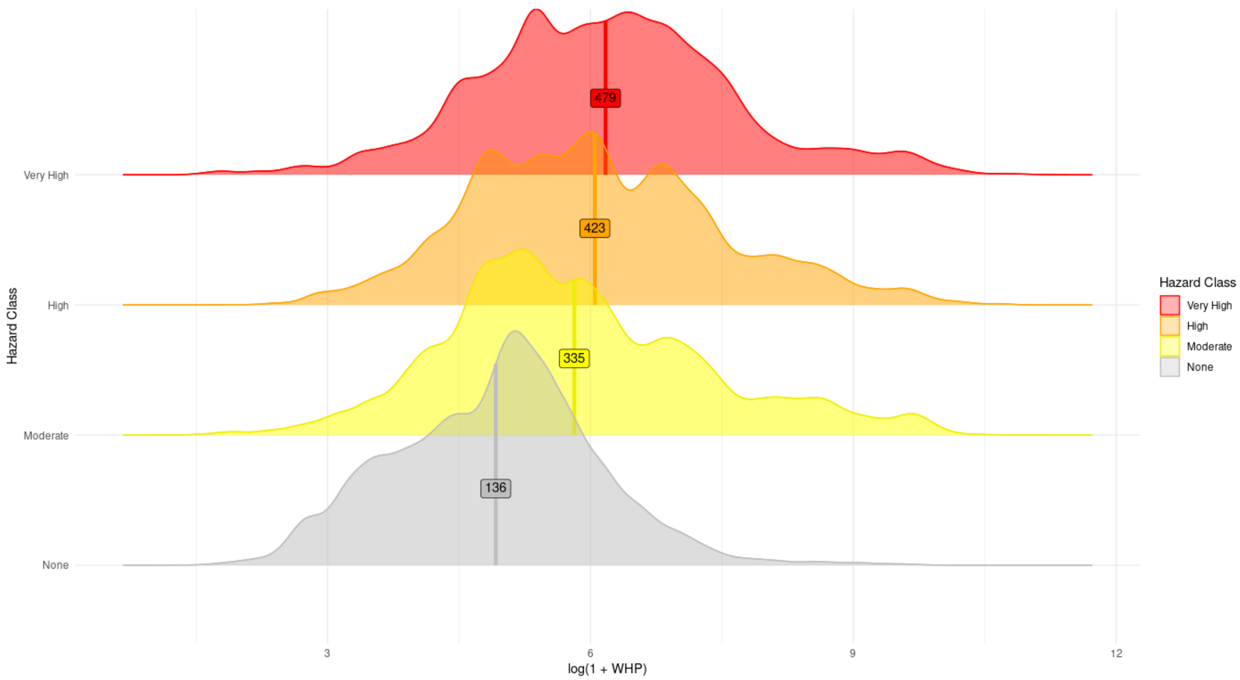
Note: This table re-estimates the model from Table 4 but limits the estimation sample to Local Responsibility Areas (LRAs) only.

Table 10: Robustness—Preference Estimates That Remove Homes Built After 2007

	Base	Q2	Q3
Disclosure	1,211.23	150.40	276.51
	495.69	13.40	14.11
log(1+WHP)	6.77	-21.69	-38.67
	60.31	3.42	3.99
Disclosure × log(1+WHP)	-194.60	-34.91	-66.07
	80.48	4.57	5.15
log(Lot size)	3,333.96	33.73	80.88
	175.78	7.23	7.69
Avg. percent meeting standard	14,705.74	1,260.58	2,080.03
	2,223.23	35.44	37.63
Percent White non-Hispanic	7,917.79	344.75	908.79
	1,367.12	22.72	27.88
log(Distance to protected areas)	-2,567.70	-5.69	-60.09
	456.37	5.34	5.52
Southern California	-	4.71	-131.13
	-	12.01	12.61
Previous fire within 5 years	-1,192.26	135.38	160.86
	812.93	29.79	32.01
MC (β_{FMC})	-2.4695	0.0043	
Individuals ($\sum_t N_t$)	23,979		
Choices ($\sum_t J_t$)	7,107		

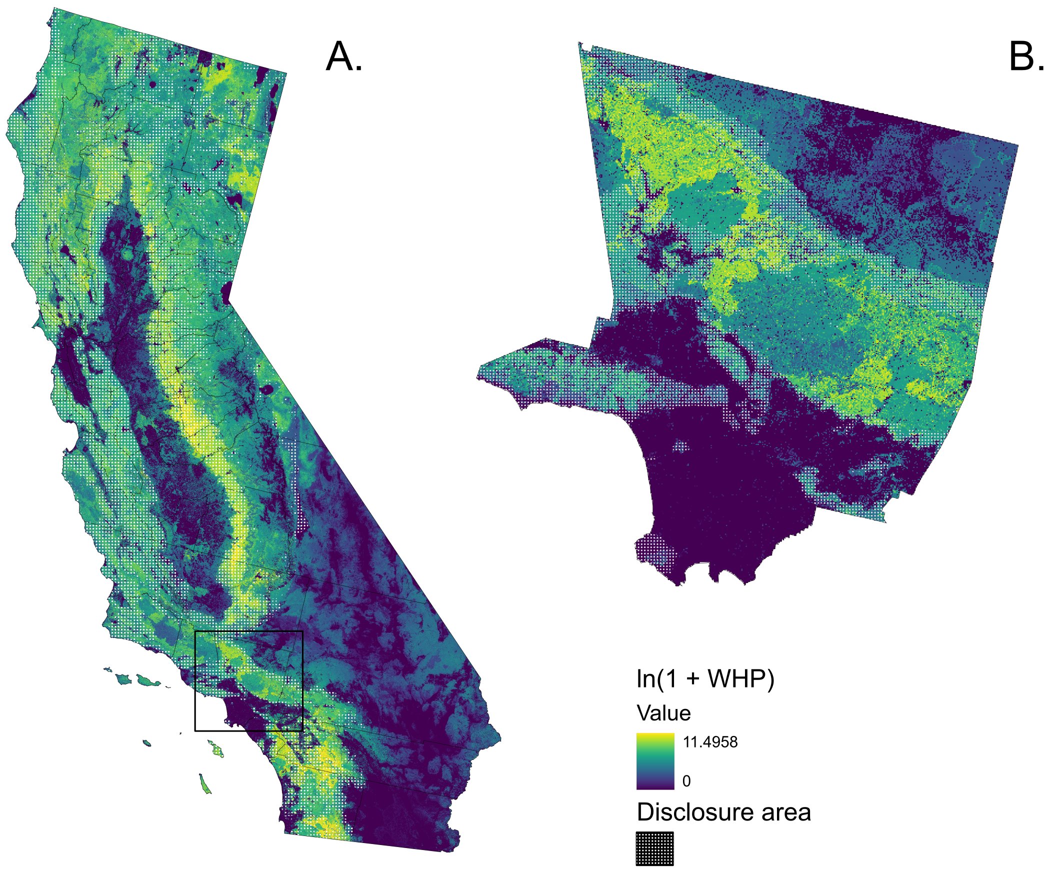
Note: This table re-estimates the model from Table 4 but removes homes built in 2008 or later from the estimation sample.

Figure 1: Distributions of Wildfire Hazard Potential (WHP) Within LRA and SRA Rates, by Hazard Class



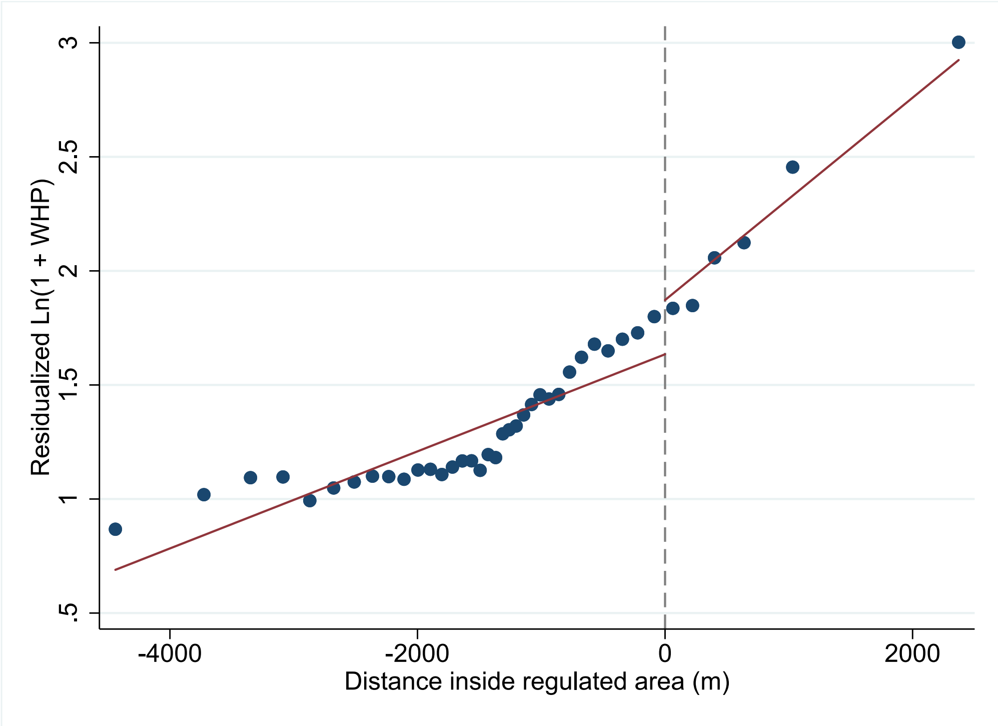
Note: To better illustrate distribution of WHP, the figure omits the large number of properties where it equals zero. The hazard class for homes in LRAs not classified as very-high FHSZs is assigned based on the hazard rating within CAL FIRE's "Nine-class" data set (described in Section 3).

Figure 2: Wildfire Hazard Potential (WHP) and Wildfire Hazard Disclosure Areas in California (panel A) and Los Angeles County (panel B)



Note: Areas with mandatory wildfire hazard disclosure are shown with white stippling. These areas comprise moderate, high, and very-high Fire Hazard Severity Zones (FHSZs) within State Responsibility Areas and very-high FHSZs within Local Responsibility Areas.

Figure 3: Continuity of Wildfire Hazard Across the Regulatory Boundary



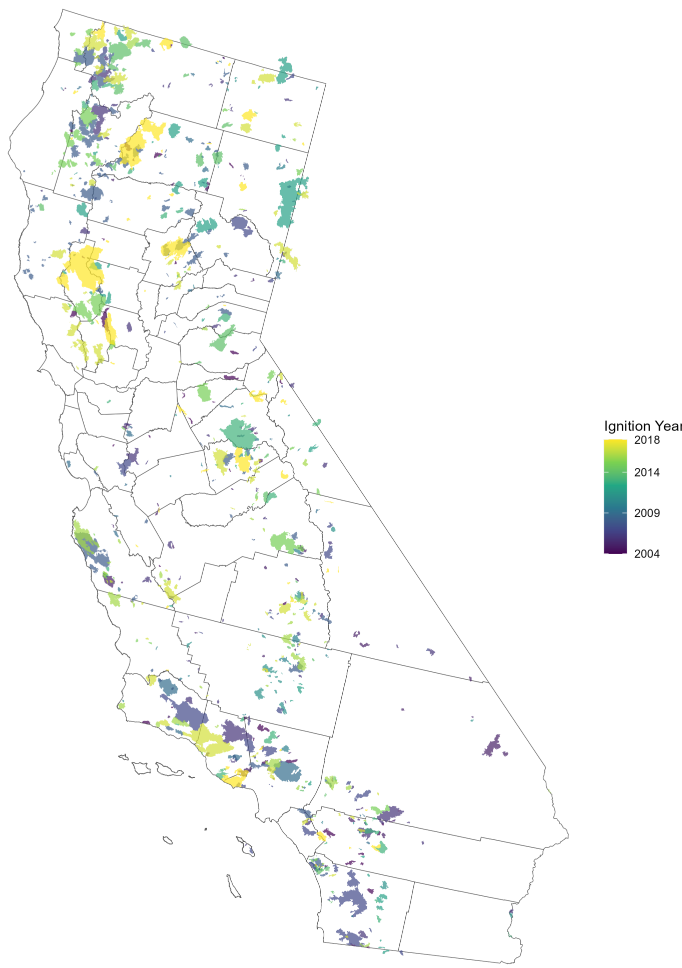
Note: Residualized $\ln(1 + WHP)$ is $\ln(1 + WHP)$ after residualizing on boundary ID fixed effects. The plot shows scatter plot of average residualized $\ln(1 + WHP)$ within 40 equally sized bins of distance to regulated area boundary. The line shown is fitted on the raw (rather than binned) data.

A Additional Tables & Figures

Table A.1: All Stage 2 Estimates for Various Model Specifications

	(1)	(2)	(3)
Specification:	Table 4	Table 5	Table 6
Disclosure	1,073**	810.3	1,073**
	(487.7)	(503.0)	(487.7)
log(1+WHP)	61.99		63.76
	(59.42)		(59.42)
Disclosure × log(1+WHP)	-257.6***		-260.4***
	(78.39)		(78.39)
Low WHP (0 < WHP < 401)	85.87		
		(247.2)	
High WHP (WHP ≥ 401)		438.1	
		(509.1)	
Disclosure × Low WHP		-547.8	
		(394.3)	
Disclosure × High WHP		-1,413**	
		(655.1)	
log(Lot size)	3,562***	3,528***	3,559***
	(175.8)	(175.3)	(175.8)
Avg. percent meeting standard	12,645***	12,852***	12,630***
	(2,213)	(2,215)	(2,213)
Percent White non-Hispanic	8,464***	8,578***	8,445***
	(1,355)	(1,356)	(1,355)
Bedrooms	-367.8**	-363.0**	-368.0**
	(154.7)	(154.8)	(154.7)
Age of Home	-32.05***	-32.45***	-32.03***
	(7.815)	(7.823)	(7.815)
log(Building area)	22,583***	22,596***	22,585***
	(465.9)	(466.1)	(465.9)
Previous fire within 5 years	-1,129	-1,201	-1,151
	(772.9)	(773.6)	(772.9)
log(Distance to protected areas)	-1,897***	-1,872***	-1,899***
	(446.8)	(447.2)	(446.8)
Incorporated area	1,249**	1,267**	1,248**
	(531.2)	(531.5)	(531.2)
Built after 2008	2,170***	2,182***	2,168***
	(496.1)	(494.7)	(496.1)
Moderate Fire Hazard Severity Zone	132.6	156.6	131.5
	(261.9)	(262.4)	(261.9)
High Fire Hazard Severity Zone	739.2***	776.4***	738.2***
	(267.6)	(265.4)	(267.6)
Very High Fire Hazard Severity Zone	-87.57	110.0	-88.60
	(504.9)	(499.8)	(504.9)
Constant	-126,756***	-127,332***	-126,732***
	(5,512)	(5,515)	(5,512)
Observations	12,104	12,104	12,104

Figure A.1: California Fires by Year, 2004–2018



B Additional Details About Datasets

B.1 Zillow ZTRAX Transaction and Assessors' Data

ZTRAX is a real estate database made available to researchers by the online real estate marketplace Zillow. It contains property characteristics data—sourced from county assessors' data—for homes across more than 2,750 US counties and data on more than 20 years of real estate transactions, including sales. We processed data from ZTRAX using information from and according to best practices outlined in Nolte et al. (2023).[26]

ZTRAX assessors' data contains variables describing homes, lots, and other real estate characteristics. These data are split across tables in the database, with some tables containing fields describing buildings (eg. number of bedrooms) and others containing fields describing the parcel (e.g., lot acreage). Parcels can be linked to more than one building. For multiple buildings in a single lot, we first filter on completeness; that is, we assume the primary building is the one with complete entries for number of bedrooms or square footage if other buildings are missing these entries. If multiple building records have complete entries, we assume the largest of the buildings (in square footage) is the primary SFR and assign its properties to the parcel as hedonic attributes.[27,28] We also remove transactions that involve large numbers of parcels under the same transaction ID to avoid including large-scale residential purchases unlikely to be associated with a single agent's moving decision. As the coordinate reference system for the latitude and longitude is not consistent across rows in this dataset, we follow Nolte et al. (2023) and transform coordinates based on the apparent majority coordinate reference system by county as reported in Section 3.2 of that work. We remove observations with missing values for latitude and longitude.

[26] Additional information on best practices can be found at `https://placeslab.org/hedonic-data-practices/`

[27] An additional issue within ZTRAX data, which the structure of the database does not account for, is that buildings may occasionally be linked to multiple parcels. This can occur, for example, when buildings cross municipal boundaries, especially with different property tax incidence in those neighboring areas. In these cases, building features may appear in ZTRAX to be identical buildings on neighboring parcels or a single building with an adjacent empty lot. We remove duplicate assessment and transaction records and parcels with completely empty records; in general, these filters should remove duplicate records of this type.

[28] Addresses (which ZTRAX uses to estimate spatial coordinates) can also sometimes be linked to multiple parcels. For addresses with multiple parcels, we use the most-current assessor data for each parcel ID, along with the most-complete assessor record to choose a parcel–location link. We remove different parcel row IDs if they share an exact latitude and longitude and avoid issues arising from this problem.

Transactions in ZTRAX include all events in which a property deed is recorded to have changed ownership, including events that are not direct sales and represent inheritances, defaults, and other nonmarket transfers. We remove all such non-arms-length transactions, as their listed sale prices do not necessarily represent market value for the property. We identify them based on the "IntraFamilyTransfer" flag available within ZTRAX transaction data. We remove rows where the reported sale price is missing or less than $5,000. Because we focus on personal residences, we keep only properties ZTRAX classifies as single family residential, rural residence, or inferred single family residential.

Finally, loan amount is an important variable for our analyses; however, the ZTRAX data set contains transactions for novel loan originations and a variety of other transaction types, including deed transfers. To determine if that loan amount is associated with a novel loan origination, we filter on the "data class ID," which categorizes transaction information based on what broad class of document (related to mortgage, deed transfer, etc.) the record comes from. We consider loans associated with either a data class ID of "both a deed transfer and a mortgage" (H) or "a mortgage" (M) to represent loan originations, as opposed to those associated with solely data from a deed transfer and likely to be remaining loan balances, which are coded "D." We remove transactions with missing loan amounts.

B.2 Home Mortgage Disclosure Act Data

Compiled by the Consumer Financial Protection Bureau (CFPB), the Housing Mortgage Disclosure Act (HMDA) Loan Application Register (LAR) contains data on demographics for individuals covering all property-backed loans in the United States.[29] This dataset was initially proposed to shine a light into potential discrimination and allow public scrutiny of lending decisions. The data are anonymized—neither the lender nor the lendee are directly identifiable through the HMDA LAR alone, and each loan is geographically identified at the purchased home's tract level. Lenders are tracked through an anonymous alphanumeric

[29] The HMDA LAR data exclude most FSA loans, which are subsidized loans often used to provide operating capital or purchase new farmland and not mandatory to report to credit agencies. As a result, ranches and other residences on agricultural lands may not show up in our dataset.

code[30] that matches to a local lending group or institution. Loan amounts are reported in 1,000-dollar buckets until 2019, when CFPB began using 10,000-dollar increments.

We are primarily interested in decisions for homeowners within California, so we filter the HMDA to focus on loans used for new property purchases in California. We filter out data on loans for upgrading homes, buying vacant land for development, or other non-property usages. Additionally, data are filtered to rows where the lender completed the loan origination process, and lendees were not rejected for the loan at any stage.

C Merging Transactions to HMDA Demographics

We follow Bayer et al. (2016) in linking our ZTRAX transaction data set to the HMDA LAR data set in order to identify buyer demographics. Data from HMDA LAR and ZTRAX transactions do not share a common identifier that allows them to be easily merged. Instead, we take advantage of the fact that both sets of data contain complete records of mortgage loans, and we identify unique loan amount matches within the same census tract and year. Specifically, we modify the procedure outlined in Billings (2019) to produce a dataset of Zillow properties with matched relevant demographic characteristics sourced from the HMDA reports.

C.1 Merging procedure

Before beginning the merging process, we clean ZTRAX and HMDA to facilitate matching and limit each data set to a set of observations we hope to produce matches for.

ZTRAX and HMDA capture different aspects of financial transactions. ZTRAX is a comprehensive record of all home sales and deed transfers. However, it includes transactions such as all-cash sales and nonsale property transfers (e.g., foreclosures, inheritances) that are not in the HMDA dataset. Conversely, HMDA records all property-backed loans, including those that do not involve a property transfer, such as

[30] These alphanumeric codes can be created by concatenating agency code, representing to whom the loan is sent (e.g., Federal Reserve System, National Credit Union Agency, Federal Deposit Insurance Commission) and respondent ID (an agency-provided identifier for the lender) for 2000 and onward, or the unified Legal Entity Identifier (LEI) from 2018 to 2023.

cash-out refinancing or remodel loans. To address this, the HMDA records are isolated to those associated with owner-occupied housing, or single-family residences where the stated loan purpose includes purchase.[31] To prepare the ZTRAX data for matching, we transform loan amounts to match the HMDA reporting style. Prior to 2018, loan amounts in the ZTRAX data are rounded to the nearest 1,000; for 2018 and after, loan amounts are assigned to the midpoint of the 1,0000-dollar range in which they fall.[32] The merging procedure proceeds as follows:

1. We identify records in the HMDA and ZTRAX data that comprise unique matches on census tract, year, and loan amount.

2. ZTRAX data contain a variable identifying lender names, but HMDA identifies lenders using unique alphanumeric identifiers.[33] Based on merged observations from Step 1, we construct a dictionary that translates lender identifiers into names. We can then identify records in the HMDA and ZTRAX data that comprise unique matches on census tract, year, loan amount, and lender. This method, building upon that in Billings (2019), allows us to find more matches than previously possible.[34]

3. Finally, we again look for unique matches on census tract, year, loan amount, and lender name. We employ the Soundex algorithm (US Social Security Administration, 1969), implemented in the stringdist package (van der Loo, 2014), to convert lender names in ZTRAX into a phonetic format. Soundex, originally used for historical record identification, is effective because it disregards nonmeaningful characters, such as punctuation, allowing us to match phonetically identical names despite potential spelling or white space discrepancies.

[31] In later years, the HMDA has data with multiple purposes per loan: if any of those purposes are for home purchase, we keep the record.

[32] For example, a loan for 523,230 dollars sold in 2015 would be rounded to 523,000. In 2020, that same loan would be assigned to the 525,000 loan amount bin. This 2020 bucketing strategy can result in large differences in actual loan amount and the reported amount; in the extreme case, a loan of 520,001 dollars in 2020 would be assigned a loan value of 525,000.

[33] Lenders in HMDA are identified by concatenating the "respondent ID" with the agency code. Post-2017, these identifiers are unified into a legal entity identifier (LEI).

[34] For changes in identifiers in 2017, the ARID 2017 table from the FFIEC is used to ensure consistency in our dictionary. Further details can be found at `https://ffiec.cfpb.gov/documentation/faq/identifiers-faq`.

C.2 Matching Results

Of the ZTRAX transactions we hope to identify demographic data for, we find 31 percent of them are matched in the HMDA LAR dataset 2009–2020. A benchmark for the performance of our matching algorithm is the number of matched transactions reported in Billings (2019). We use that study because we base much of our procedure on their work and they, unlike other similar work focusing on California, use ZTRAX as their source of transaction data to match with the HMDA LAR rather than the commonly-used CoreLogic dataset.

We compare our total number of matched observations to that reported in Billings (2019), rather than comparing match rates, because of differences in our studies in data preparation and target samples. Because our studies rely on the same data source, we expect that identical approaches over the same time frame should produce the same number of matches. Billings (2019) report 254,355 HMDA-ZTRAX matched pairs in California in 2007–2016.[35] In comparison, over same time frame, our modified matching algorithm produces 750,461 ZTRAX-HMDA matches. The improvements we see relative to Billings (2019) are due to the additions of Steps 2 and 3 in the algorithm they described.[36]

D Financial Moving Costs

We construct measures of household-specific financial moving costs and fire experience by using buyer and seller names to identify the location households resided prior to moving. To identify the neighborhood a household is leaving, we match buyer names in our merged ZTRAX–HMDA dataset with seller names from property sales California-wide in the two years before or after the purchase date. Consequently, our dataset only includes financial

[35] Billings (2019) do not publish state-level match counts, but they can be deduced for California based on their Figure 8 and Table 1.

[36] Billings (2019) report a 45 percent match rate for California; however, this match rate is not directly comparable to the one reported in our paper. Their rate is calculated based on matching 254,355 of 566,393 ZTRAX loan transactions from California 2007–2016. In contrast, our procedure matches 750,461 out of 2,280,841 loan transactions during the same period. Because Billings (2019) lacks a detailed description of the sample filtering steps taken to limit the sample to the 566,000 transaction observations from which they match, we cannot directly compare match rates. However, the increase in the total *number* of matches produced from the same raw data indicates increased performance of our matching algorithm.

moving costs for households relocating within California; we are not able to identify previous locations for homeowners moving from other states or first-time homeowners.

ZTRAX provides buyer and seller names, but these are inconsistently reported, sometimes including middle names or initials and occasionally just first initials without full first names. Therefore, before beginning the matching algorithm, we remove white space and punctuation from all names and convert them to uppercase.

Our procedure for matching for buyers to their origin home is as follows:

1. For each property in the HMDA-matched data set, we identify exact matches between buyer and seller names for properties that sold within two years of the focal home's purchase date. When buyers have sold multiple properties in the four-year window, we assume they are moving from their most recently sold property.

2. For transactions without a match in the initial search, we look for exact matches based on first and last names, omitting middle names.

3. Finally, for unmatched buyer or seller names, or those with only initials in the first or middle name fields, we match based on a string composed of the first initial, middle initial(s), and full last name.

In each step, we require unique matches between buyer names and the name of a single seller within a two-year window on either side of the purchase date; this conservative choice reduces the number of unique matches between buyer and seller names but also false positive matches. We limit the sample to those buyers who sold a property within a maximum of one year before their purchase, which helps to restrict attention to transactions that occurred connected to a move, rather than, for example, purchase of an investment property or second home. We match 39 percent of HMDA-matched buyers in ZTRAX home purchase transactions to the location they moved from.