Sorting Over Wildfire Risk Information*

Sarah Bass[‡] Michael Fogarty[§]

Job Market Paper

This version: September 23, 2026

Click here for the most recent version

Abstract

We study how information about climate risk changes household beliefs, housing demand, and residential sorting in equilibrium. We embed a Bayesian model of belief updating into a structural model of neighborhood choice in which households have heterogeneous preferences over housing prices and wildfire risk. We use the introduction of a wildfire risk disclosure policy in California as an information shock that changes information about wildfire risk without changing the underlying physical hazard, allowing us to identify how belief updating affects location choices and housing prices in equilibrium. Our estimates suggest that prior to the disclosure policy, households exhibit substantial uncertainty in their beliefs about neighborhood wildfire risk. We find that the divergence in beliefs about neighborhood risk induced by the disclosure policy is equivalent to the effect of exposure to about 8 wildfire events. Counterfactual simulations that remove the disclosure policy show that this change in beliefs induces substantial re-sorting of households based on their relative levels of sensitivity to wildfire risk: more risk-averse, mortgage-financed buyers shift away from areas newly perceived as higher risk, while less risk-averse buyers, particularly cash buyers and those purchasing homes for non-primary use, sort in. Re-sorting is also supported by general-equilibrium price effects: prices in neighborhoods newly perceived as risky fall relative to neighborhoods believed to be safer. As a result, disclosure reallocates wildfire exposure across households. Our findings show that the provision of wildfire risk information can reshape housing markets and the distribution of climate risk, even when the underlying physical hazard remains unchanged.

*We would like to thank our dissertation committee advisors Jesse Gregory, Corina Mommaerts, Martin O'Connell, and Jeff Smith. We also thank Ben Collier, Chao Fu, JF Houde, Laura Schechter, Chris Taber, Chris Timmins, and Serena Xu for their comments on this work.

‡Bass: Terry College of Business, University of Georgia; Email: sarahbass@uga.edu

§Fogarty: Department of Economics, University of Wisconsin; Email: michael.fogarty@wisc.edu

1 Introduction

The allocation of climate risk across households depends not only on the spatial distribution of physical hazards, but also on households' beliefs about those hazards when making location decisions. When households have incomplete or biased beliefs about risk, providing new information can change where they choose to live, inducing re-sorting across space and changes in equilibrium housing prices. These equilibrium responses can, in turn, change which households ultimately bear climate risk. Understanding how households trade off wildfire risk against other amenities, and how providing risk information alters households' decisions about where to live, is important to understanding which households will bear climate risks. In this paper, we study how information about wildfire risk updates household beliefs and how these belief changes reshape residential sorting and the equilibrium allocation of risk across households.

In 2020, California enacted legislation that requires sellers to disclose wildfire risk information to homebuyers during the home sales process. The policy introduced standardized information about wildfire exposure during the home buying process without changing the underlying physical risk. In this paper, we use this policy reform to examine how climate risk information affects housing market outcomes and residential sorting. We find that households re-sort in response to the risk information policy. As households update their beliefs about wildfire risk, more risk-sensitive buyers sort away from disclosed areas, while less risk-sensitive and more price-sensitive buyers sort in.

A growing literature shows that households often underestimate exposure to climate risks such as flooding and wildfire, but update their beliefs in response to new information (Bakkensen and Barrage, 2022; Gibson and Mullins, 2020; Weill, 2026). Consistent with this idea, prior work finds that wildfire risk disclosure in California affects housing market outcomes, reducing prices in high-risk areas and shifting the composition of buyers away from mortgage-financed, owner-occupant purchases toward cash buyers and second-home investors, particularly in areas with more severe insurance constraints (Bass and Fogarty, 2025; Ma et al., 2024). Together, these findings suggest that disclosure induces substantial re-sorting across wildfire risk. However, reduced-form evidence cannot separately recover the beliefs and preferences that generate these sorting responses. In particular, it cannot quantify the extent to which households misperceive wildfire risk or how correcting those beliefs interacts with heterogeneous preferences for risk and housing prices to generate re-sorting.

More broadly, reduced-form estimates of disclosure effects are inherently local and do not capture how responses propagate through the broader housing market. By design, these estimates compare areas near disclosure boundaries where underlying wildfire risk is similar, limiting their ability to predict how information would affect sorting patterns across the full distribution of risk or under alternative policy regimes. In addition, they do not recover the underlying preferences governing household location choices or account for how equilibrium price adjustments across neighborhoods amplify or dampen sorting responses. Understanding how information reshapes residential sorting therefore requires a framework that jointly models household preferences, beliefs about risk, and equilibrium housing prices.

This paper develops and estimates a structural model of neighborhood sorting in which households have heterogeneous preferences over housing prices and wildfire risk and beliefs about wildfire risk that update in response to new information. We use spatial variation in wildfire risk disclosure to identify households' prior beliefs about wildfire risk and their heterogeneous preferences for risk and housing prices. The model allows us to quantify the extent to which households misperceive wildfire risk in the absence of disclosure and to evaluate how belief updating and equilibrium price adjustments jointly determine sorting patterns in response to climate risk information.

The model builds on the empirical patterns documented in Bass and Fogarty (2025) by allowing preferences to differ across buyer types defined by financing status and intended home use (primary residence vs. non-primary use). In the absence of disclosure, buyers have beliefs about wildfire risk that are uncertain and potentially inaccurate. The disclosure policy induces households to update their beliefs: neighborhoods subject to risk disclosure are perceived to be riskier, while neighborhoods not subject to disclosure are perceived to be safer. These changes in perceived risk shift housing demand across neighborhoods and generate equilibrium price adjustments.

We estimate the model using microdata on the universe of single family home transactions in California from 2017-2024 alongside a panel of neighborhood characteristics that includes prices, wildfire risk, observed market shares, and measures of insurance market conditions. We first construct neighborhood housing price indices using a hedonic model that combines rich property characteristics with machine learning methods. We then estimate heterogeneous household preferences for housing prices and wildfire risk using simulated maximum likelihood, leveraging observed neighborhood choices to recover how preferences and beliefs differ across buyers. To address the endogeneity of housing prices to local amenities, we instrument for neighborhood prices using a shift-share approach that combines cross-neighborhood differences in buyer type preferences with national trends in the composition buyer types.

Estimating the model reveals substantial heterogeneity in how buyers trade off housing prices and wildfire risk. Cash buyers are significantly more sensitive to housing prices than mortgage-financed buyers, while risk sensitivity is greatest among mortgage-financed buyers purchasing primary residences and lowest among cash and non-primary purchasers. We also find that households have substantial uncertainty surrounding their prior beliefs, and this uncertainty leads households to update substantially in response to the disclosure policy. Our estimates suggest that disclosure drives a divergence in beliefs about wildfire risk between disclosed and non-disclosed neighborhoods equivalent to the exposure to 8 wildfire events.[1]

To quantify the impacts of the disclosure policy on sorting across buyer types and equilibrium prices, we conduct a counterfactual exercise in which we shut off the disclosure policy in the post period (2021-2023), holding the rest of the estimated parameters fixed. Under the counterfactual distribution of beliefs, we find a substantial re-allocation of buyer types across neighborhoods. Holding housing supply fixed, the policy raises the share of the most risk-tolerant buyer type (cash buyers of secondary properties) in disclosed neighborhoods by 5.8 percentage points in 2021, while lowering the share of mortgaged buyers of primary homes by 7 percentage points. We see smaller effects in the opposite direction in neighborhoods not subject to the disclosure policy. The magnitude of the equilibrium price effects are contingent on the absolute value of buyers' distaste for wildfire risk, while the sorting results only rely on the cross-type differences. As a result, the implied price changes across disclosed and non-disclosed neighborhoods varies from -0.2% to -10.7% across different estimates. These effect sizes bookend the 4% price effect we estimate in Bass and Fogarty (2025), so we view our range of estimates as bounds on the actual price effect.

A growing literature studies how climate risk information enters housing markets through disclosure policies, insurance pricing, and platform-based information provision. Empirical work shows that formal disclosures and information shocks—such as wildfire hazard disclosures (Ma et al., 2024), regulatory flood maps (Hino and Burke, 2021), and large-scale platform experiments (Fairweather et al., 2023)—affect housing search behavior, transaction outcomes, and equilibrium prices, although capitalization is often incomplete or context-dependent, consistent with shifting salience (Bin and Landry, 2013). Recent

[1]It is important to recognize that the typical wildfire event is relatively small: in our sample of wildfires across the state of California spanning 2012-2023, the median wildfire burned 36 acres, while the mean burned about 3,500. The distribution of wildfire sizes is heavily right-skewed, with the largest 10% of fires accounting for over 97% of acres burned.

work expands beyond price effects to examine additional margins of adjustment, including financing and insurance: Bass and Fogarty (2025) show that wildfire risk disclosure shifts transactions toward cash purchases and reduces owner-occupancy, while Collier et al. (2023) documents substantial declines in insurance take-up following premium increases. While this literature establishes that climate risk information affects housing market behavior, less is known about the household beliefs and preferences that generate these responses or how they aggregate in equilibrium. We recover these underlying beliefs and preferences and use them to characterize how information provision reshapes housing demand and residential sorting across the broader market.

A second literature focuses on how individuals perceive climate risk and how those perceptions shape economic behavior, emphasizing belief formation, updating, and heterogeneity. Early work shows that direct experience with extreme events can update perceived risk, as reflected in housing market outcomes (Kousky, 2010), while more recent studies document that risk perceptions respond to both informational signals (e.g., flood maps) and lived experience (Gibson and Mullins, 2020). A central finding is that beliefs are often systematically biased or misinformed: (Mulder, 2024) shows that incorrect risk classification can distort insurance and adaptation decisions with welfare consequences, and (Bakkensen and Barrage, 2022) documents substantial heterogeneity and underestimation of flood risk that helps explain mixed evidence on price capitalization. Complementary experimental evidence finds that information provision can shift stated beliefs about climate risk but does not always translate into behavioral change (Deryugina and Shurchkov, 2016). We build on this work by recovering households' perceptions of wildfire risk directly from revealed residential choices. The shock to beliefs induced by the disclosure policy allows us to disentangle beliefs from heterogeneity in households' preferences for wildfire risk as separate channels that impact household residential choices.

A third strand of literature examines how climate risk shapes residential sorting and migration decisions, building on a foundational literature estimating equilibrium models of location choice (Epple and Sieg, 1999; Epple et al., 2001; Bayer and Timmins, 2007) and heterogeneous preferences for neighborhood attributes (Bayer et al., 2007; Ioannides and Zabel, 2008). Applied work incorporates climate risk into these frameworks, showing that households sort across locations based on exposure to environmental risk and that these responses are highly heterogeneous: Bakkensen and Ma (2020) find differential sorting over flood risk by income and race with important distributional implications. We extend this literature by allowing the risk households perceive when choosing where to live to differ from underlying physical risk and to change with new information. Incorporating belief updating into an equilibrium model of residential sorting allows information itself to reshape the spatial allocation of households, even when the underlying distribution of physical risk is unchanged.

We proceed as follows. We begin with institutional background describing the rollout of the disclosure requirement and the information contained in the natural hazard disclosure form. We next introduce the model of belief updating in response to information provision, and embed households' beliefs about wildfire risk in to a model of neighborhood choice. We then turn to describe the data we use to estimate the model. Next, we describe the estimation procedure and identification of key parameters and present the accompanying results. We close with a discussion of implications, limitations, and conclusions.

2 Institutional Background

Wildfire risk has become an increasingly important consideration for households in California, where annual wildfire burned area increased roughly fivefold between 1972 and 2018 (Williams et al., 2019). Households choosing where to live must weigh the benefits of fire-prone locations against the risks of property damage, financial loss, and disruption. These risks vary substantially across space, making wildfire exposure a key dimension along which neighborhoods differ.

Households can respond to wildfire risk along several margins, including investments in property-level mitigation, insurance coverage, and residential location. These responses are shaped by both private and public institutions. Building standards can substantially reduce wildfire losses, including through spillovers that reduce structure-to-structure fire spread (Baylis and Boomhower, 2026), while public wildfire suppression can subsidize development in high-risk areas by shifting some of the costs of wildfire protection to the government (Baylis and Boomhower, 2023). The financial consequences of wildfire also extend beyond direct property losses: wildfire damage can disrupt mortgage repayment and expose gaps between insurance settlements and rebuilding costs (Biswas et al., 2023). Insurance markets provide another important mechanism for allocating wildfire risk. Insurers increasingly use granular measures of wildfire exposure to classify and price properties (Boomhower et al., 2024; Boomhower and Fowlie, 2026), while mortgage lenders can respond to residual disaster risk by changing the terms on which households can finance homes in risky areas (Sastry, 2026). These institutional responses affect both the costs households face from living in high-risk locations and which households are able or willing to bear those costs.

Despite the importance of wildfire risk, households may not fully perceive the risks they face when making these decisions. Limited information about both wildfire exposure and the returns to self-protection can inhibit private adaptation (Boomhower et al., 2023; Baylis and Boomhower, 2026). Moreover, insurance prices may not effectively communicate underlying risk to households. Mulder and Kousky (2023) show that risk-based insurance pricing may have limited effects on adaptation when the risk information underlying those prices is not directly communicated to households. Information frictions are particularly important for residential location decisions because purchasing a home is a long-lived investment that determines a household's exposure to wildfire risk, insurance costs, and potential property losses.

California's wildfire risk disclosure policy introduced standardized information about wildfire exposure into housing transactions. Beginning in 2021, California added new fields to its preexisting natural hazard disclosure form indicating whether a property is located in a designated wildfire hazard zone based on Fire Hazard Severity Zone (FHSZ) maps created by the California Department of Forestry and Fire Prevention.[2] Properties located in areas designated by CalFIRE as "Very High" fire hazard severity zones, as well as areas designated as "High" or "Moderate" that lie within state-managed fire protection zones, are required to check "Yes" on the relevant wildfire fields in the Natural Hazard Disclosure Form.

The information provided in the disclosure is coarse, indicating exposure to mapped risk zones rather than property-specific risk, and includes a general notice that buyers may face challenges obtaining homeowners insurance. The disclosure does not change the underlying risk, but provides buyers with new information about that risk at the time of purchase. This distinction between underlying risk and information about risk is central to our empirical setting: the policy changes the information available to households without directly changing the physical hazard they face.

3 Model

To study how wildfire risk information disclosure shapes residential sorting, we develop a discrete choice model of neighborhood demand in which households have beliefs about wildfire risk, and the risk

[2] Figure A1 displays the updated Natural Hazard Disclosure form that includes the new wildfire hazard fields

disclosure policy induces a Bayesian updating of households' beliefs about wildfire risk.

We adapt the belief updating model from Barahona et al. (2023), who model the impact of a food labeling policy on consumer demand for breakfast cereals, to better fit our setting. In our context, wildfire risk disclosure is a noisy, but still informative, signal of underlying wildfire risk. This causes households to shade down the extent to which they update their beliefs compared to the fully informative case.

In our setup, wildfire risk disclosure ($D_j \in \{0, 1\}$) is a noisy but informative signal of underlying wildfire risk. Upon signal receipt, households update their beliefs in a Bayesian manner, with the extent of the updating moderated by the amount of noise in the disclosure signal.

The motivation for this approach comes from Bass and Fogarty (2025), which documents that the disclosure policy not only reduced home prices in disclosed areas, but also shifted the composition of buyers toward cash purchases and away from primary home use. These changes in buyer composition are difficult to explain through a full-information model, and are instead consistent with an information story in which disclosure makes wildfire risk more salient at the point of purchase, causing different types of buyers to update their behavior differently. This paper directly models that information channel by treating disclosure as a trigger that induces households to update their beliefs about wildfire risk.

We close the model with a housing supply specification that governs how changes in demand are split between price adjustments and quantity responses. In equilibrium, neighborhood prices adjust until supply equals demand in each neighborhood. The remainder of this section describes the demand and supply sides of the model in turn.

3.1 Neighborhood Demand

We propose a neighborhood choice model that treats neighborhoods as differentiated products, as in Bayer et al. (2007). Households buying a house in period t solve the following discrete choice problem:

$$\max_{j \in J_t} V_{ijt} \tag{1}$$

The choice set J_t consists of all neighborhoods defined by the intersection of ZCTAs with the risk disclosure area, plus an outside option of relocating outside of California. Equation 2 gives the indirect utility of household i choosing neighborhood j in period t.

$$V_{ijt} = \theta^P_{z(i)} P_{jt} + \theta^R_{z(i)} E_{it}[r_j|D_j] + X_{jt}\beta + \xi_{jt} + \varepsilon_{ijt} \tag{2}$$

P_{jt} is a price index for the price of housing services in neighborhood j in year t, $E_t[r_j|D_j]$ is the mean of household i's belief about wildfire risk in neighborhood j in period t, and X_{jt} are observable neighborhood characteristics. ξ_{jt} is an unobserved (to the econometrician) neighborhood amenity. ε_{ijt} is an iid taste shock distributed according to the Type-1 extreme value distribution. We normalize the scale of the taste shocks by setting $\sigma_\varepsilon = 1$.

Price sensitivity varies across different buyer types z, but is constant over time in the following way:

$$\theta^P_{z(i)} = \bar{\theta}^P + \sum_{k=1}^{3} z_{ik}\theta^P_k$$

Here, z_{ik} is a set of indicators for whether household i belongs to types $k \in \{1, 2, 3\}$. Likewise, risk aversion also varies by buyer types z, but is consistent with respect to wildfire risk beliefs $E_t[r_j|D_j]$,

$$\theta^R_{z(i)} = \bar{\theta}^R + \sum_{k=1}^{3} z_{ik}\theta^R_k$$

Observable household heterogeneity follows buyer types Z, which are given by the intersection of two characteristics: mortgage financing status (cash versus mortgage financed purchase) and home purchase type (owner-occupied versus non-owner-occupied). This intersection creates four types of buyers, which are detailed in Table 1. We treat mortgage-financed buyers of primary homes as the reference category for the heterogeneous preferences over price and risk because they are the modal buyer type, accounting for an outright majority of home purchases in the state of California.

3.1.1 Prior Beliefs

Before the wildfire risk disclosure policy comes into effect in 2021, households have a prior belief about wildfire risk in neighborhood j that are a function of R_j, a continuous measure of the underlying wildfire risk in neighborhood j, as well as F_{jt}, a cumulative measure of realized wildfire exposure that evolves with observed wildfire activity across the state. We include realized fire activity as a determinant of prior beliefs because a substantial literature documents that direct exposure to natural hazard events shifts households subjective risk assessments (Kousky, 2010).

$$r_j \sim N(\lambda_R R_j + \lambda_F F_{jt}, \sigma_r^2)$$

The parameters λ_R and λ_F reflect the relative loadings of the continuous risk measure and cumulative wildfire exposure onto beliefs r_j, while σ_r^2 reflects the degree of *uncertainty* in households' beliefs about wildfire risks. While σ_r^2 does not directly enter the utility function through the scalar summary $E_t[r_j|D_j]$, this parameter plays a critical role in governing the extent to which households update their beliefs about r_j in response to the risk disclosure signal D_j.

3.1.2 Disclosure Policy

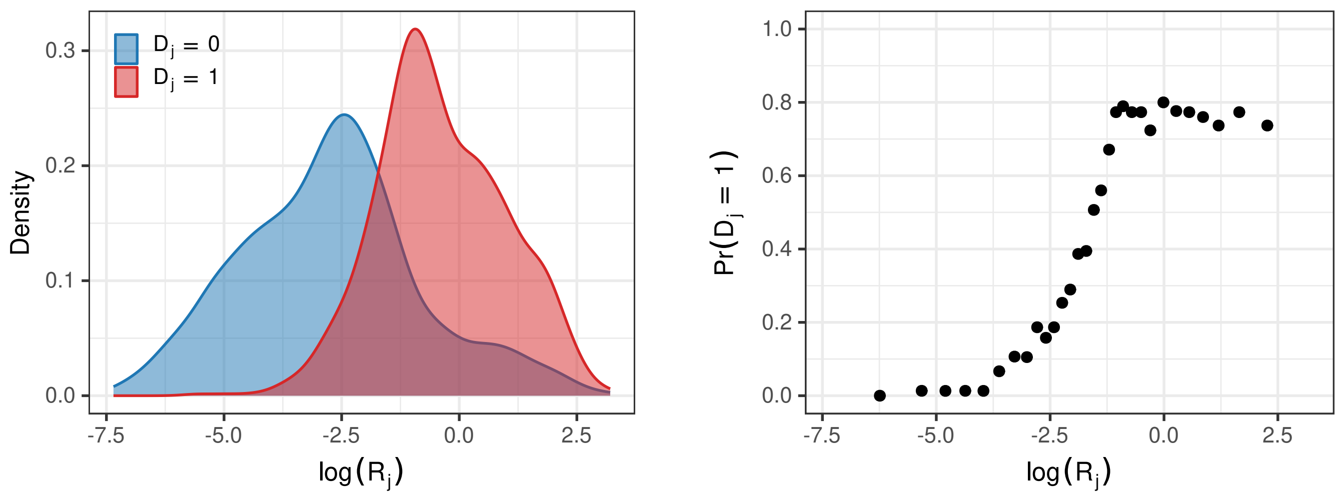
One key difference between our setting and the Barahona et al. (2023) cereal labeling paper the binary wildfire risk disclosure signal is not perfectly informative of the underlying risk level. In this setting D_j is a noisy, but informative, signal of underlying risk. The CalFIRE hazard zones do not perfectly segment neighborhoods by observable risk score; Figure 1a shows the overlap between the distribution of risk scores for disclosed (red) and non-disclosed (blue) neighborhoods.

Figure 1b shows a binned scatter plot of the conditional probability that a neighborhood receives a disclosure designation as a function of the risk score R_j. After a sharp run up, the conditional probability flattens off at around 0.8, suggesting that even some very high risk neighborhoods are not given the disclosure flag.

We propose a simple model of a noisy disclosure policy rule, modifying a standard Probit with a ceiling probability π that captures the fact that even high risk neighborhoods go undisclosed.

$$\Pr(D_j = 1) = \pi \times \Phi\left(\frac{R_j - \overline{r}}{\tau}\right)$$

Here, $\{\pi, \overline{r}, tau\}$ the government's disclosure "policy function". Following the model above, we can externally estimate the policy parameters using maximum likelihood to recover the ceiling probability, cutoff, and noise parameters, and take these parameters as given in the model of households' belief updating



(a) Risk scores by disclosure status

(b) Probability of disclosure as a function of risk

Figure 1: Wildfire Risk and Disclosure Status

Notes. The unit of observation is the neighborhood, the intersection of a ZCTA with the FHSZ-regulated area; $D_j = 1$ denotes neighborhoods subject to the disclosure requirement from 2021. The horizontal axis in both panels is the log of the continuous risk index R_j, the mean of the USFS 2023 wildfire hazard potential raster over the neighborhood polygon divided by 1,000. Panel (a) plots kernel density estimates of $\log R_j$ separately for disclosed and non-disclosed neighborhoods. Panel (b) groups neighborhoods into 30 equal-count bins of $\log R_j$ and plots, for each bin, the share of neighborhoods with $D_j = 1$ against the bin's mean $\log R_j$. The 83 neighborhoods with a hazard score of zero, for which $\log R_j$ is undefined, are excluded, leaving 2,263 neighborhoods (989 disclosed, 1,274 not disclosed).

process.

The implication of the noise in the policy rule is that it introduces additional uncertainty into the updating process and makes the disclosure signal less informative, so as a result households will shade their updating relative to the perfectly informative signal case.

3.1.3 Belief Updating

After observing D_j, households update their beliefs r_j and form a new expectation $E_t[r_j|D_j]$. There are two cases, depending on which disclosure regime neighborhood j belongs to. Conditional on receiving the disclosure $D_j = 1$, households expectation of the risk score will increase. Likewise, the posterior will shift lower in response to receiving the signal $D_j = 0$.

The formula for the mean of the posterior in each case is as follows:

$$E_t[r_j|D_{jt}=1] = \mu_{jt} + \frac{\sigma_\eta^2}{\sqrt{\sigma_\eta^2+\tau^2}} \frac{\phi(\hat{\alpha}_j)}{1-\Phi(\hat{\alpha}_j)} \tag{3}$$

$$E_t[r_j|D_{jt}=0] = \mu_{jt} - \frac{\sigma_\eta^2}{\sqrt{\sigma_\eta^2+\tau^2}} \frac{\phi(\hat{\alpha}_j)}{\Phi(\hat{\alpha}_j)} \tag{4}$$

$$\hat{\alpha}_j = \frac{\overline{r}-\mu_{jt}}{\sqrt{\sigma_\eta^2+\tau^2}} \tag{5}$$

Here, μ_{jt} is the mean of the prior distribution. Each updating formula consists of the prior mean, plus or minus a scaling factor times a belief updating factor. The scaling factor is the signal to total noise ratio, and shades down the size of the update as the disclosure rule gets noisier. As τ^2, the noise in the policy rule, increases, households shade down the extent of their belief updating; as σ_r^2, the uncertainty in the prior, increases, households shade down their updating less and the scaling factor approaches 1.

The updating quantity itself is an inverse Mills ratio term that takes as its argument a variance-standardized distance between the prior mean and the disclosure policy threshold. As Barahona et al. (2023), the impact of disclosure on the posterior mean is non-linear in the distance between the prior and the cutoff. Focusing on the $D_j = 1$ case, when the prior $\mu_{jt} < \bar{r}$, the size of the update increases in the distance from the disclosure threshold $\bar{r}$, while beliefs update relatively little when $\mu_{jt} \geqslant \bar{r}$, as the signal corroborates the households' prior belief.

Figure 2 illustrates the belief updating process with a calibrate example. The rows of the figure represent neighborhoods with relatively low and high prior beliefs about wildfire risk, and the columns represent $D = 0$ and $D = 1$. The posterior means with the noisy disclosure rule are smooth and not perfectly truncated as in the no-noise case. Directionally however, the effects move in a similar manner. Households have relatively small updates to their beliefs when the disclosure signal matches the direction of the prior (above or below the estimated cutoff threshold). However, they update substantially in response to a contradictory signal: when μ_{jt} is low and $D = 1$, the posterior mean jumps considerably and vice-versa.

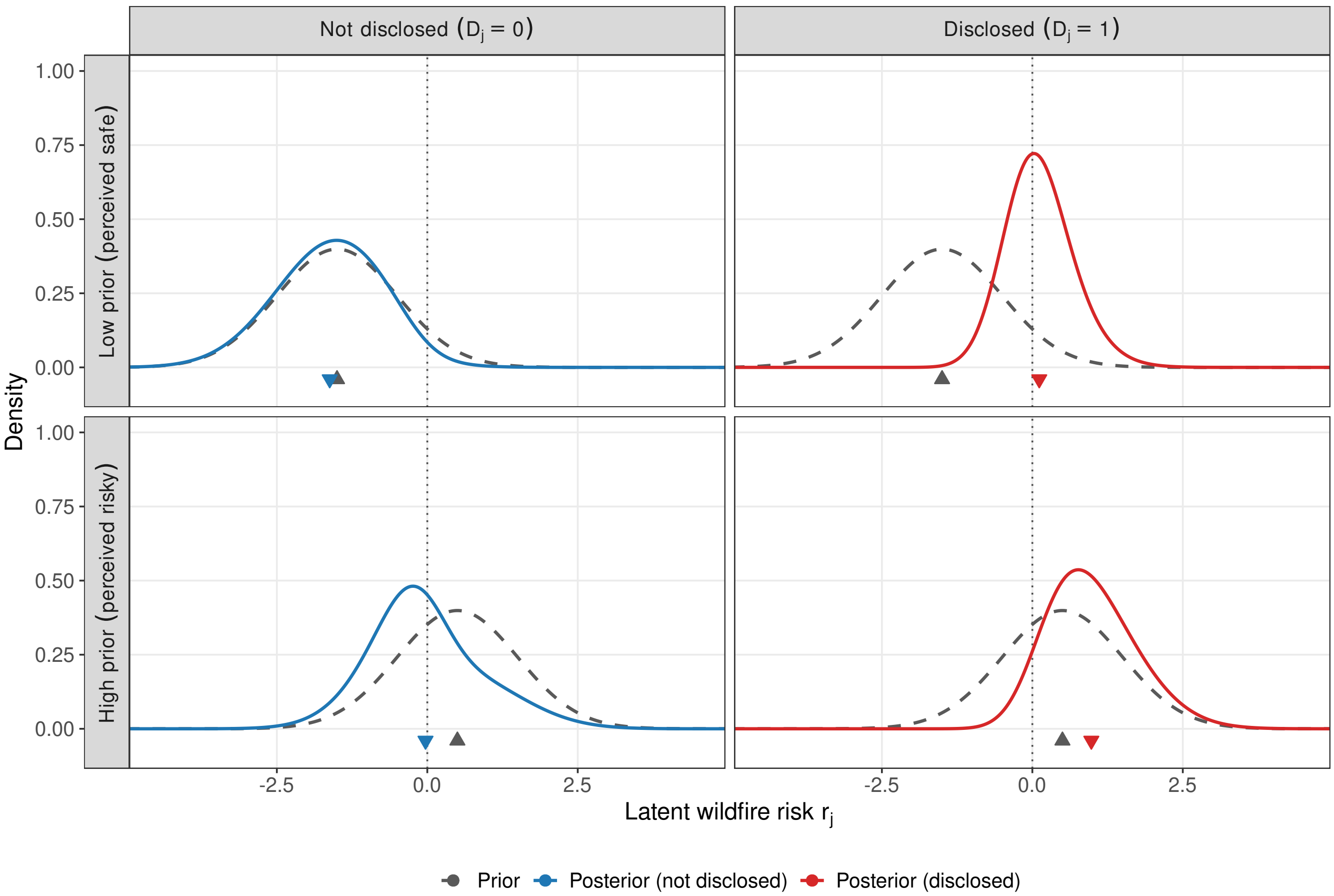


Figure 2: Illustrative Example of Belief Updating Process

Notes. Illustrative prior and posterior beliefs about latent wildfire risk r_j under the ceiling-aware updating rule, evaluated at example parameter values rather than the estimates: $\sigma = 1$, $\tau = 0.5$, $\pi = 0.78$, and $\bar{r} = 0$. Rows vary the prior mean, $\mu_{jt} = -1.5$ (low prior) and $\mu_{jt} = 0.5$ (high prior); columns condition on disclosure status. In each panel the dashed gray curve is the prior $r_j \sim N(\mu_{jt}, \sigma^2)$, and the solid curve is the posterior, which is proportional to the prior times the probability of the observed disclosure status under the rule $\Pr(D_j = 1 \mid r_j) = \pi\,\Phi((r_j - \bar{r})/\tau)$. After a disclosure the ceiling π cancels, so it affects only the non-disclosed column, where the likelihood is $1 - \pi\,\Phi((r_j - \bar{r})/\tau)$. Posteriors are normalized numerically. Triangles on the horizontal axis mark the prior mean (upward) and the posterior mean (downward), and the dotted vertical line marks the threshold $\bar{r}$.

3.1.4 Choice Probabilities and Neighborhood Demand

Given a set of parameters Θ, household i's probability of choosing neighborhood j in period t follows the standard logit form:

$$\pi_{ijt} = \frac{\exp\left(\theta^P_{z(i)} P_{jt} + \theta^R_{z(i)} E_{it}[r_j|D_j] + X_{jt}\beta + \xi_{jt}\right)}{1 + \sum_{k\in J_t} \exp\left(\theta^P_{z(i)} P_{kt} + \theta^R_{z(i)} E_{it}[r_k|D_k] + X_{kt}\beta + \xi_{kt}\right)} \tag{6}$$

The model-implied market share for neighborhood j in year t is found by aggregating individual choice probabilities across all households:

$$\sigma_{jt}(\theta, \delta) = \frac{1}{N_t} \sum_{i=1}^{N_t} \pi_{ijt} \tag{7}$$

A key step in the estimation procedure is matching the model-implied market shares σ_{jt} to the observed market shares s_{jt} constructed from the data. As shown in Berry et al. (1995), for a given parameter vector θ there exists a unique vector of mean utilities δ that equates model-implied and observed market shares. This inversion delivers $\delta_{jt}(\theta)$ as a function of the preference parameters, which we describe further in Section 5.

3.2 Housing Supply

To close the model and conduct counterfactual policy analyses, we need to specify the supply side of the housing market. The supply specification determines how changes in neighborhood demand induced by the disclosure policy are split between price adjustments and quantity responses. We consider three specifications, each of which embeds different assumptions about the short- and long-run responsiveness of housing supply.

The first specification assumes perfectly inelastic supply. In the short run, housing is durable and, especially in California, new construction is difficult due to permitting constraints and land scarcity. Under inelastic supply, all demand-induced shifts are fully capitalized into neighborhood prices, with no quantity response. This is the most restrictive assumption and is most appropriate for studying the short-run price and sorting effects of the disclosure policy.

The second specification assumes iso-elastic supply, with a finite supply elasticity taken from the literature (Baum-Snow and Han, 2024; Saiz, 2010, e.g.,). This allows for longer-run supply responses to demand shocks, where price increases in a neighborhood attract new construction and price declines reduce it. The iso-elastic specification is better suited for longer-run counterfactual analyses that allow for supply-side readjustment to changes in demand induced by different counterfactual policy experiments.

The third specification follows Glaeser and Gyourko (2005), who argue that housing supply is asymmetric due to the durability of the existing housing stock. Under this specification, supply is inelastic in response to negative demand shocks, since existing homes cannot be costlessly demolished, but has a positive elasticity in response to positive demand shocks, where new construction is feasible. This hybrid specification captures the asymmetric nature of housing supply responses and the idea that, even in the long run, durable housing stocks will magnify the price impacts of negative demand shocks.

4 Data

We study the housing market in California from 2017-2024. This eight-year time span gives us a balanced four years of coverage before and after the wildfire risk disclosure policy was implemented in 2021.

Estimating the demand model requires two primary data inputs, which we describe in turn. First, we construct a repeated cross section of the neighborhood choices and characteristics of single family home buyers in the state of California for our sample period. We also build a panel of "product characteristics" for each neighborhood that comprises the choice set for households in the model.

For the purposes of bringing the model to data, we define neighborhoods as the geographic intersection of Census Zip Code Tabulation Areas (ZCTAs) with the area subject to the risk disclosure. The state of California is divided into about 1,700 ZCTAs.[3] After taking the spatial intersection of the ZCTAs with the area subject to the disclosure policy, we are left with about 2,300 neighborhoods.

[3]The set of ZCTAs do not fully cover the state of California - there are swathes of largely uninhabited Federal land that are not covered by any ZCTA. We implicitly exclude these areas from households' choice set.

4.1 Household Choices and Characteristics

We construct a dataset of household's neighborhood choices and characteristics from a repeated cross section of home transactions derived from the CoreLogic OwnerTransfer data file. Our sample captures all arms-length transactions of single family homes in the state of California from 2017-2024.

4.1.1 Neighborhood Choice

We code household neighborhood choice by geocoding each transaction to our neighborhood definitions using parcel-level latitude and longitude from the CoreLogic data. By geocoding to a fixed GIS representation of the neighborhoods, we ensure that neighborhood choice outcomes are measured consistently across all years in our dataset, even though ZCTAs were updated by the Census Bureau for the 2020 Census. We use the 2010 ZCTAs for the entire sample period.

4.1.2 Buyer Type Information

Households in the model feature multiple types of heterogeneity, some of which is observable to the econometrician and some of which is latent or unobservable.

Households vary in an observable way along two dimensions: cash/mortgage financing status, and intended usage (primary vs. non-primary property usage). Both buyer financing and intended primary home usage are directly observable in the CoreLogic data. We define four observable buyer types as the intersection of these two binary characteristics. Table 1 shows the distribution of the four different buyer types in the data. The modal buyer is mortgage-financed and intends to purchase the home for primary use, so we set these buyers as the baseline or left-out category in the model. Overall, about 76% of buyers in our sample purchase a home with mortgage financing, and 64% of home purchases are intended as the buyer's primary residence. We cannot distinguish between different secondary uses, such as second homes, short-term rentals or long-term rentals, so we choose to focus on the primary vs. non-primary margin.

4.2 Neighborhood Characteristics

The other primary data input for the model is a panel of neighborhood characteristics. We construct a yearly panel that captures the "product characteristics" of each neighborhood in our dataset for each year

Table 1: Buyer Type Distribution

Type	Financing	Purchase Type	Frequency
0	Mortgage	Primary	0.56
1	Mortgage	Non-Primary	0.20
2	Cash	Primary	0.08
3	Cash	Non-Primary	0.17
N		3,916,856	

Notes:
This table reports the distribution of buyer types in the CoreLogic transaction data, covering all arms-length sales of single-family homes in California from 2017–2024. Buyers are categorized by financing status and intended use. The majority of transactions are mortgage-financed purchases of primary residences. Cash buyers are disproportionately represented in non-primary (e.g., second home or investment) purchases, highlighting systematic differences in buyer composition across financing types.

from 2017 through 2024. For each neighborhood-year, we measure several characteristics: a home price index P_{jt}, a continuous index of wildfire risk R_j, realized wildfire events F_{jt}, market share s_{jt}, disclosure status $Disclosed_j$, a measure of insurance market tightness, and two measures of home insurance prices. The home price index, market share, disclosure status, and insurance conditions all vary at the neighborhood-year level, while wildfire risk is treated as time-invariant given that the underlying physical hazard does not change materially over our sample period. Table 2 provides summary statistics for the key neighborhood characteristics, for the entire sample as well as comparing neighborhoods subject to disclosure ($D = 1$) and those not subject to risk disclosure ($D = 0$). We describe the source and construction of each of these measures in turn.

4.2.1 Home Price Indices

Estimating the model requires neighborhood-level price indices that vary at the yearly level. We construct our own price indices using the universe of single family home transactions in California from CoreLogic. We follow Sieg et al. (2002) and estimate neighborhood price indices as neighborhood-year fixed effects in a hedonic regression of the following form:

$$\ln p_{kt} = a_{j(k),t} + H_k\Gamma + u_{kt}$$

Here, p_{kt} is the observed transaction price for house k sold in neighborhood $j(k)$ in year t. H_k is a vector of house characteristics that includes lot size, home square footage, the number of bedrooms and bathrooms,

Table 2: Summary Statistics: Neighborhood Characteristics

	Overall	$D_j = 1$	$D_j = 0$
Panel A: Prices and market shares			
Log price index	13.2	13.1	13.2
	(0.501)	(0.567)	(0.447)
Price level ($1,000s)	595	605	587
	(642)	(924)	(314)
Market share ($\times 10^{-3}$)	0.449	0.193	0.630
	(0.552)	(0.282)	(0.620)
Panel B: Wildfire risk			
Wildfire hazard potential R_j	0.950	1.68	0.435
	(2.18)	(2.60)	(1.63)
Cumulative fire-years	0.923	1.52	0.499
	(1.61)	(1.91)	(1.18)
Any prior fire (%)	38.0	58.6	23.4
	(48.5)	(49.3)	(42.4)
Panel C: Insurance market conditions			
Private premium ($/policy-year)	1,420	1,578	1,308
	(885)	(811)	(918)
FAIR premium ($/policy-year)	1,515	1,933	1,200
	(1,250)	(1,267)	(1,140)
FAIR share of policies (%)	4.33	7.09	2.38
	(9.34)	(12.0)	(6.18)
Neighborhood-years	15,579	6,461	9,118
Neighborhoods	2,339	994	1,345
ZIP codes	1,713		

Notes. Means with standard deviations in parentheses. Unit of observation is the neighborhood-year, where a neighborhood is the intersection of a ZCTA with the FHSZ-regulated area; $D_j = 1$ denotes the regulated portion, subject to the disclosure requirement from 2021. The price index P_{jt} is the ZCTA $\times$ FHSZ $\times$ year fixed effect from the hedonic regression, re-centered at the mean log sale price. Market share is the neighborhood's share of all moves in the state-year. Wildfire hazard potential R_j, the continuous risk index, is the mean of the USFS 2023 continuous WHP raster over the neighborhood polygon, divided by 1,000. Cumulative fire-years counts distinct years in which a CAL FIRE perimeter of at least one acre intersected the neighborhood, over fire years 2012 to $t-1$. Insurance variables are earned premium per unit of earned exposure from the CDI Community Service Statements, which end in 2022 and are reported at ZIP level only. The FAIR premium is observed for 91.8% of neighborhood-years, and is missing where FAIR exposure is zero.

and more.

We link the sample of transactions from the CoreLogic OwnerTransfer file with detailed structure and property characteristics from the CoreLogic Property file to create our estimation sample, which spans about 3.8 million transactions over our eight year sample period.

To more flexibly estimate the hedonic model, we generalize the linear model above into the following form:

$$\ln p_{kt} = a_{j(k),t} + f(H_k) + e_{kt}$$

To estimate the model, we proceed in two steps, applying the logic of the Frisch-Waugh-Lovell theorem to first flexibly estimate the effects of the structural characteristics H_k on prices p_{kt} using a random forest model (Wright and Ziegler, 2017) to estimate the hedonic function $f(\cdot)$.

$$\ln p_{kt} = f(H_k) + u_{kt}$$

Then, having estimated the function $\hat{f}(\cdot)$ via random forest, we construct the first-step residuals:

$$\hat{u}_{kt} = \ln p_{kt} - \hat{f}(H_k)$$

To recover the zip-year fixed effects, we then run the following regression:

$$\hat{u}_{kt} = a_{j(k),t} + e_{kt}$$

Out of the box, the random forest already substantially outperforms the linear model, and after tuning the model hyperparameters, model performance (as assessed by RMSE or R^2) increases further.

A further advantage of the random forest approach is that it neatly handles missing data in the predictors by treating missings as a distinct, qualitatively different category than non-missing predictors. This approach is both more flexible and introduces less estimation error than imputation-based methods of addressing missing predictor variables.

Figure 3 compares the estimated neighborhood price indices from the random forest model and the linear hedonic model. The estimates from the random forest model are less strongly associated with the mean neighborhood sales price (the x-axis variable), which suggests that the random forest model does a better job of disentangling the relationship between underlying neighborhood quality (as measured by the price index) and observable home characteristics.

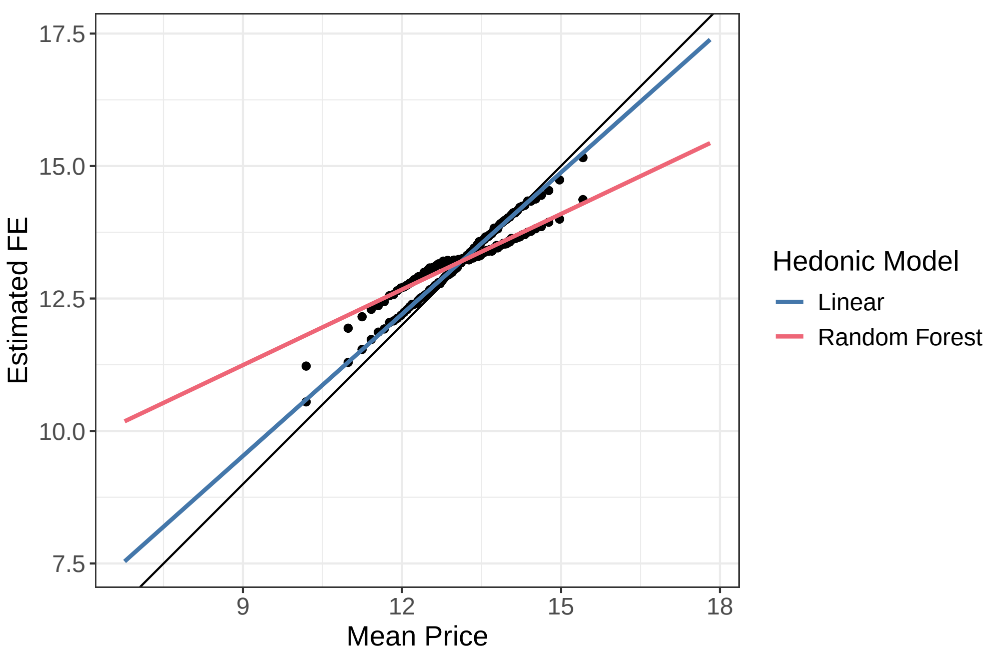


Figure 3: Comparison of Estimated Neighborhood Price Indices

Notes:
The black 45-degree line corresponds to mean observed neighborhood prices. The blue line shows price indices estimated with a linear hedonic model, while the red line shows estimates from a random forest model. Both approaches adjust for observable housing characteristics, but the random forest does so more flexibly. The greater divergence of the random forest estimates from the 45-degree line indicates that it more effectively removes the influence of compositional differences in housing stock (e.g., larger or higher-quality homes in higher-priced neighborhoods), yielding a cleaner measure of underlying neighborhood value.

4.2.2 Wildfire Risk and Realized Exposure

Wildfire risk in California exhibits a substantial degree of spatial concentration. Our primary measure of wildfire risk is the U.S. Forest Service's continuous Wildfire Hazard Potential (WHP) index. We also compare this continuous, forward-looking risk score to two other measures of wildfire risk: historical wildfire perimeters and discrete Fire Hazard Severity Zones (FHSZs) developed by the California Department of Forestry and Fire Protection (Cal Fire). Figure 4 maps each measure of wildfire risk side-by-side.

The Cal Fire FHSZ maps divide the state into four discrete fire risk categories: Low, Moderate, High, and Very High. In addition to serving as a publicly-available source of fire risk information, the FHSZs were used by the state legislature to define the areas where sellers are legally required to disclose wildfire risk at the time of sale.

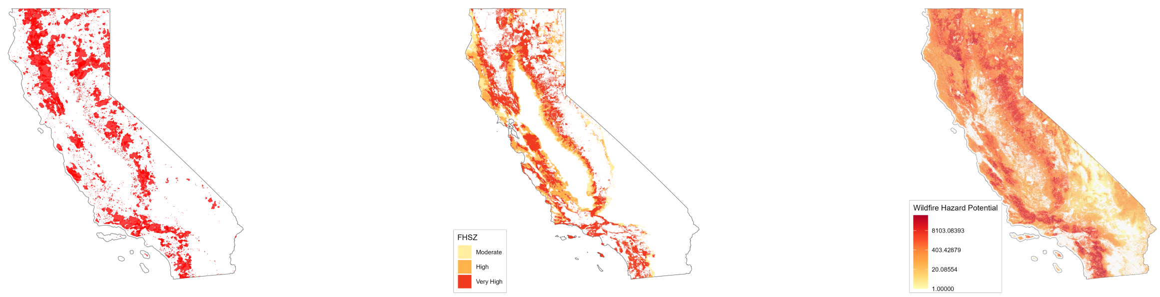
The U.S. Forest Service WHP dataset provides a spatially detailed assessment of wildfire risk, quantifying the likelihood and intensity of wildfire activity at a fine resolution of 270 square meters. This measure accounts for factors such as vegetation, topography, historical fire occurrence, and modeled fire behavior to produce a comprehensive risk assessment. The WHP measure is scaled from 0-10,000, where zero is associated with little to no wildfire hazard potential (examples of areas with a WHP score of zero are a densely developed urban neighborhood or an empty desert, both of which lack sufficient vegetation to fuel a wildfire).

Using a continuous measure of wildfire risk allows us to capture the substantial heterogeneity in wildfire risk within the discrete Fire Hazard Severity Zones that were used to determine the areas subject to the wildfire risk disclosure policy.

We also construct a measure of realized wildfire activity at the neighborhood-year level. We geocode annual fire perimeter data from the CAL Fire Fire and Resource Assessment Program (FRAP) to our ZCTA-Disclosure neighborhood definition. We define a fire exposure event as any wildfire activity with an area greater than 1 acre occurring within neighborhood j in year t.[4]

We then calculate a cumulative measure of wildfire exposure. For neighborhood j in year t, F_{jt} is the count of years in which the neighborhood had any level of fire exposure through year $t-1$, starting from 2012. This starts each neighborhood with a five-year history of wildfire exposure, and exposure counts

[4] This 1-acre threshold removes 8.5% of wildfire events but only alters about 2% of neighborhood-year fire exposure measures.



(a) Wildfire Perimeters **(b)** Fire Hazard Severity Zones **(c)** USFS Wildfire Hazard Potential

Figure 4: Wildfire exposure and risk in California.

Notes:
Panel (a) shows historical wildfire perimeters.
Panel (b) shows Fire Hazard Severity Zones (FHSZ) used to define disclosure policy boundaries.
Panel (c) shows the continuous Wildfire Hazard Potential (WHP) measure.

Each property is linked to these maps via geolocation to assign both continuous risk and disclosure status. All three measures identify similar high-risk regions at a broad geographic level. However, the FHSZ boundaries impose discrete cutoffs on what is fundamentally a continuous risk landscape, meaning that nearby properties can face similar underlying risk but differ in disclosure status due to boundary definitions.

continue to evolve across our sample period as fire events occur. The distribution of wildfire exposure is heavily zero-inflated - almost 2/3 of neighborhoods never experience a wildfire event over our sample period.

4.2.3 Market Shares

Matching the model-implied market shares σ_{jt} to observed market shares s_{jt} is a key step in the model estimation procedure and delivers the vector of mean utilities, δ_{jt}.

We construct market shares using the same sample of arms-length purchases of single-family homes that forms the basis for our buyer sample. Let Y_{ijt} be an indicator that takes the value one if buyer i purchases a home in neighborhood j in year t. We define the market shares s_{jt} in the following way:

$$s_{jt} = \frac{1}{N_t} \sum_{i=1}^{N_t} Y_{ijt}$$

Where N_t is the number of transactions in our CoreLogic sample for year t, plus the estimated net outflow of homebuyers from the state of California in year t. To estimate the outflow in each year, we take net domestic migration estimates from the Census Population Estimates Program and multiply the net migration by the owner share of out-movers estimated from the ACS PUMS.[5] Both the owner share of out-movers and raw migration increase over time, leading to an increase in the share of households choosing the outside option throughout the sample period. Figure A2 plots the time series of estimated out-migration of homeowners from 2011-2025, as well as the implied outside option share.

4.2.4 Disclosure Status

California's wildfire risk disclosure requirement was signed into law in 2020 and took effect in 2021. The law requires sellers of residential properties in designated high-risk fire zones to provide buyers with a standardized written disclosure of the parcel's wildfire risk at the time of sale. The geographic scope of the requirement is defined by Cal Fire's Fire Hazard Severity Zone (FHSZ) maps: properties classified as High or Very High fire hazard severity are subject to mandatory disclosure.

We construct a neighborhood-level disclosure indicator $Disclosed_j$ that equals one for neighborhoods located within the disclosure area in years after the policy took effect (2021 onward), and zero otherwise. Neighborhoods that lie entirely outside the High and Very High FHSZs are never disclosed; neighborhoods that lie entirely within the disclosure area are disclosed from 2021 forward. Because we define neighborhoods as the spatial intersection of ZCTAs with the FHSZ boundary, neighborhoods are either fully inside or fully outside the disclosure area by construction – a ZCTA that straddles the boundary is split into two separate neighborhoods, one disclosed and one undisclosed. This split adds about 600 neighborhoods to our choice set compared to a ZCTA-only neighborhood definition.

4.2.5 Insurance Market Conditions

The market for homeowners insurance in California is facing pressures from changes in underlying risk of fires and strict regulation on insurance premiums. During our study period, several major insurers stopped writing new policies in the state, highlighting the strain in the homeowners insurance market.

We use data from the California Department of Insurance's Community Service Statement (CSS) to provide supplementary information about the status of the local homeowners insurance market. We use annual data from the CSS covering 2017-2022 to calculate annual ZIP code level coverage share of the California FAIR Plan, as well as two measures of insurance prices. The FAIR Plan is California's

[5] We impute the 2020 owner share as the mean of the 2019 and 2021 shares due to data interruptions caused by COVID-19

state insurer of last resort, which offers more expensive and less comprehensive coverage than private insurers. FAIR coverage excludes certain perils covered under standard homeowners' policies, such as theft, earthquakes, and liability coverage.[6] A higher share of FAIR Plan policies in a ZIP code signals that private insurance is scarce or prohibitively expensive.

The CSS provides detailed information on the number of policies and amount of coverage underwritten broken down by insurance provider, line of business, and ZIP code at an annual frequency. Our primary measure of insurance market tightness is the fraction of all homeowners policies underwritten in ZIP code j in year t that are written by the FAIR Plan:

$$\text{FAIR Share}_{jt} = \frac{\text{FAIR policies}}{\text{Total homeowners policies}}$$

The FAIR Plan coverage data serves as a proxy for the share of homeowners in a given ZIP code who have experienced difficulty obtaining insurance coverage through the private insurance marketplace. Figure 5 shows that reliance on the California FAIR Plan rose sharply between 2017 and 2022, but the increases were not uniform across the state. Average ZIP-level FAIR exposure in 2017 was 2.57%, which increased to 8.26% by 2022. Figure 5 shows the level of fair plan reliance in California in both 2017 and 2022.

The largest increases in FAIR Plan reliance were concentrated in the Sierra Nevada foothills, Northern California, and parts of Southern California—areas where wildfire activity intensified. This spatial pattern mirrors areas of intensified wildfire activity and reflects private insurers' withdrawal from high-risk markets, underscoring growing insurance scarcity in high-fire-risk areas of the state. As a result, we include insurance market conditions as a separate characteristic that households value in the model.

We also calculate two ZIP-level measures of insurance prices. Both measures capture *average* insurance prices in neighborhood j in year t for a given class of homeowners insurance policy. For both privately-underwritten and FAIR plan policies, we calculate the average premium as the total earned premiums for that year divided by the toal number of policy-years in that ZIP code for that year.

[6]Homeowners who obtain coverage under the FAIR Plan typically also require supplemental coverage for these excluded hazards from a private insurer.

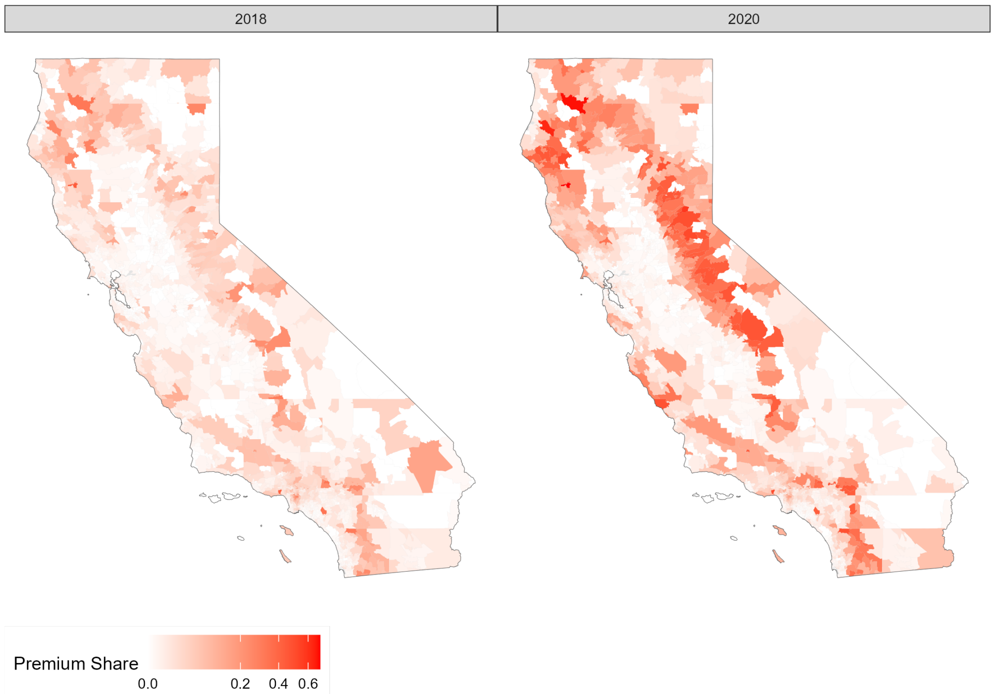


Figure 5: FAIR Plan Exposure in California

Notes:
Panel (a) 2018 share of total homeowners insurance premiums written by the California FAIR Plan.
Panel (b) 2020 share of total homeowners insurance premiums written by the California FAIR Plan.

Data come from the California Department of Insurance Community Service Statements. Between 2018 and 2020, the FAIR Plan's presence increased sharply, particularly in the Sierra foothills, Northern California, and parts of Southern California. This shift reflects growing difficulties in securing standard homeowners insurance in wildfire-prone areas. The rising reliance on the FAIR Plan indicates increasing market withdrawal by private insurers and highlights the limited availability of affordable and comprehensive coverage for households facing high environmental risk.

5 Estimation and Identification

Estimation of the demand model follows the procedures laid out in Berry et al. (1995) and Bayer and Timmins (2007) for estimating discrete-choice models of neighborhood demand. We leverage the microdata on individual-level neighborhood choices to estimate the individual-specific parameters using simulated maximum likelihood, rather than the nested GMM-IV approach used in settings where only aggregate market shares are observed.

It is convenient to separate the individual-specific and shared components of the indirect utility V_{ijt}:

$$
\begin{aligned}
V_{ijt} &= \mu_{ijt} + \delta_{jt} + \varepsilon_{ijt} \\
\mu_{ijt} &= \sum_{k=1}^{3} z_{ik} \left(\theta_k^P P_{jt} + \theta_k^R E_t[r_j|D_j]\right) \\
\delta_{jt} &= \bar{\theta}^P P_{jt} + \bar{\theta}^R E_t[r_j|D_j] + X_{jt}\beta + \xi_{jt}
\end{aligned}
$$

Here z_{ik} is an indicator for whether household i belongs to buyer type $k \in \{1, 2, 3\}$. The individual-specific component μ_{ijt} captures heterogeneity in price and risk sensitivities across types (as well as the components of the belief updating block), while the mean utility δ_{jt} captures the average valuation of neighborhood j in period t common to all buyers.

This decomposition separates the estimation problem into two main steps. In the first, we estimate the type-specific heterogeneity parameters $\{\theta_k^P, \theta_k^R\}$, the parameters governing the belief updating mechanism $\{\lambda_R, \lambda_F, \sigma\}$, and the implied mean utilities δ_{jt} jointly using simulated maximum likelihood and the BLP contraction mapping. In the second step, we decompose the mean utilities δ_{jt} into their constituent parts using an IV regression to address endogeneity between P_{jt} and ξ_{jt}. In an initial step, we estimate the disclosure policy parameters $\{\bar{r}, \tau, \pi\}$ outside of the model using maximum likelihood.

It is worth noting that the data on which we estimate the model are flows of buyers, while the demand model that we have written is fundamentally static (although we temporally segment the market by years t). For the purpose of estimating the model, we take homebuyers' selection into the decision to purchase a new home in year t as given, and use the choices of these buyers, conditional on the decision to purchase, to identify the demand parameters in the model. However, there are various dynamic features of the moving decision, such as selection into the purchasing decision, housing endowments, or the continuation value associated with different neighborhoods, that we leave unmodeled. While this is a limitation, we are still able to leverage much of the information revealed by households' choices across neighborhoods to identify household preferences for neighborhood characteristics. We discuss the identification of specific parameters in the individual-specific and common components of utility below in detail.

5.1 Estimating the Disclosure Rule

Disclosure status is assigned at the neighborhood (ZCTA $\times$ disclosure status) level and does not vary over time, so the rule is estimated on the cross-section of neighborhoods $j = 1, \ldots, J$. Let $D_j \in \{0, 1\}$ indicate that neighborhood j lies in the FHSZ-regulated area, and let $r_j > 0$ denote its wildfire hazard score. The rule is a probit in log risk, subject to a ceiling on the disclosure probability:

$$\Pr(D_j = 1 \mid R_j) = \pi \times \Phi\left(\frac{R_j - \bar{r}}{\tau}\right) \tag{8}$$

where Φ is the standard normal cdf, $\bar{r}$ is the disclosure threshold on the log-risk scale, $\tau > 0$ governs how sharply the probability rises around that threshold, and $\pi \in (0, 1]$ is the highest attainable disclosure probability. This specification nests the standard probit when $\pi = 1$.

We externally estimate the parameters $\{\pi, \bar{r}, \tau\}$ using maximum likelihood, and treat the resulting parameter estimates as the commonly-known government disclosure policy function that households take as given in the belief updating model.

5.2 Maximum Likelihood Estimation

We leverage the individual-level data on actual neighborhood choices to estimate the parameters associated with the individual-specific component of utility μ_{ijt} via simulated maximum likelihood. This approach allows us to identify these parameters using a revealed-preference logic: the differential tendency of buyer types to choose neighborhoods with different bundles of characteristics reveals their preferences for those characteristics.

5.2.1 Identification

The type-specific price sensitivities θ_k^P and risk sensitivities θ_k^R are identified from the systematic tendency of different buyer types to locate in neighborhoods with different price levels and beliefs about risk exposure: if mortgage-financed primary-home buyers are more likely than cash buyers to avoid neighborhoods they believe to be higher risk, that differential sorting pattern identifies the type-specific risk coefficients. Likewise, buyer types that tend to purchase homes in more or less expensive neighborhoods will be ascribed different levels of price sensitivity.

The parameters λ_R and λ_F represent the relative loadings of the continuous neighborhood risk measure R_j and cumulative neighborhood wildfire exposure F_{jt} into prior beliefs about neighborhood risk r_j. Because beliefs about wildfire risk have no natural scale, we require a normalization; otherwise the tuple $\left\{\theta_k^R, \lambda_F, \lambda_R\right\}$ is only identified up to a multiplicative constant c. We choose to normalize $\lambda_R = 1$, which renders θ_k^R and λ_F interpretable in log-WHP (risk score) units. Fixing this, λ_F is identified based on the relative contribution of the risk score and fire exposures that best rationalize how households sort based on their risk sensitivities as described above; this identification is aided by the cross-sectional and time series variation in F_{jt} for neighborhoods with similar risk scores R_j.

σ_r^2, which corresponds to the uncertainty in households' prior beliefs about neighborhood wildfire risk, plays a key role in governing the extent to which beliefs update in response to the disclosure policy. Higher values of σ_r^2 relative to τ^2, the noise in the disclosure policy rule, will lead households to update their beliefs more in response to the disclosure policy signals. Identification of this parameter comes from changes in household choices in the post-policy period relative to the pre-policy periods – if households behave as if disclosed neighborhoods are riskier relative to non-disclosed neighborhoods in the post period, the model with rationalize that behavior with higher values of σ_R^2

5.2.2 Estimation Procedure

Given a guess at the parameter vector $\hat{\theta} \in \Theta$, we use the contraction mapping from Berry (1994) to invert the observed market shares s_{jt} and computed the implied vector of mean utilities δ_{jt} required to make the model-implied market shares $\sigma_{jt}(\delta_{jt})$ match the market shares from the data. We start from an initial guess of δ_{jt}^0[7] and iterate the contraction mapping until convergence:

$$\delta_{jt}^{h+1} = \delta_{jt}^h + \log(s_{jt}) - \log(\sigma_{jt})$$

where h indexes iterations of the contraction mapping. Convergence occurs when $\left\|\delta_{jt}^{h+1} - \delta_{jt}^h\right\| < \epsilon$.

Given both $\hat{\theta}$ and $\delta_{jt}(\hat{\theta})$, we compute the individual choice probabilities. For household i choosing to

[7]We initialize δ^0 as the zip-year fixed effects that rationalize the observed s_{jt} under a null logistic regression model. That is, $\delta_{jt}^0 = \log\left(\frac{s_{jt}}{1-s_{jt}}\right)$

purchase in year t, the probability of selecting neighborhood j is given by:

$$\pi_{ijt} = \frac{\exp(\mu_{ijt} + \delta_{jt})}{1 + \sum_k \exp(\mu_{ikt} + \delta_{kt})}$$

Then, with the choice probabilities calculated, we can then construct the log-likelihood function:

$$\begin{aligned}
\ell(\hat{\theta}) &= \sum_t \sum_i \sum_j Y_{ijt} \log \pi_{ijt} \\
&= \sum_t \sum_i \sum_j Y_{ijt} \log(\exp(\mu_{ijt} + \delta_{jt})) - Y_{ijt} \log \left(1 + \sum_k \exp(\mu_{ikt} + \delta_{kt})\right) \\
&= \sum_t \sum_i \sum_j Y_{ijt}(\mu_{ijt} + \delta_{jt}) - Y_{ijt} \log \left(1 + \sum_k \exp(\mu_{ikt} + \delta_{kt})\right)
\end{aligned}$$

We maximize $\ell(\theta)$ using quasi-Newton methods. Standard errors are computed using the model-misspecification robust sandwich estimator (Hansen, 2022, sections 10.15, 10.19).

To aid in model convergence, we enforce a partition in which the type specific risk sensitivities θ_k^R and the prior weight on wildfire exposure λ_F are estimated solely on the pre-policy (2017-2020) years. Then, holding these parameters fixed, we estimate the belief uncertainty parameter σ_r^2 using the post-policy data. This restriction explicitly enforces the identification logic presented above. Enforcing this segmentation aids model convergence by preventing explosive behavior where estimates of σ_r^2 blow up and the θ_k^R are driven to zero. This restriction results in a trade-off of reduced estimation efficiency from estimating the model parameters on a subset of the total available data, but the estimates from this procedure are still sufficiently precise to be worth making the tradeoff.

5.3 Mean Value Decomposition

Decomposing the mean utilities δ_{jt} into their components presents an endogeneity challenge. We expect neighborhood prices P_{jt} are to be correlated with unobserved amenities ξ_{jt}: desirable neighborhoods command higher prices precisely because of qualities that are not fully captured by the observed characteristics X_{jt}. Failing to account for this correlation produces an upward-biased estimate of the mean price sensitivity $\bar{\theta}^P$.

We want to estimate the following equation (identical to the formulation of the common component of neighborhood utilities) in order to recover the mean preferences for neighborhood attributes associated with the left-out buyer category (mortgaged buyers of primary homes) as well as the neighborhood amenities:

$$\delta_{jt} = \bar{\theta}^P P_{jt} + \bar{\theta}^R E_t[r_j|D_j] + X_{jt}\beta + \varphi_j + \xi_{jt}$$

where φ_j is a neighborhood fixed effect that absorbs time-invariant unobserved amenities. We address the remaining endogeneity between P_{jt} and ξ_{jt} using an instrumental variables approach.

$\bar{\theta}^R$ is identified off of within-neighborhood variation in beliefs $E_t[r_j|D_j]$. Because the neighborhood risk scores R_j are fixed (and therefore collinear with the neighborhood fixed effects φ_j), identification stems from accumulation of fire exposure F_{jt} and the from the disclosure policy D_j.

5.3.1 Price Instruments

To isolate plausibly exogenous variation in P_{jt} that is uncorrelated with ξ_{jt}, we adapt the shift-share instrument proposed by Cook et al. (2023) to instrument for neighborhood prices to our context. The "shifts" are national (ex-California) growth rates in the share of homes purchased buy different buyer types Z_i. The "shares" are the propensities of different buyer types Z_i to purchase homes neighborhoods j, estimated using data spanning 2012-2016, before our primary sample period (Bartik, 1991; Goldsmith-Pinkham et al., 2020). We construct our instrument W_{jt} in the following way:

$$W_{jt} = \frac{\sum_z g_{zt} N_z p_{jz}}{\sum_z N_z p_{jz}}$$

Here g_{zt} is the national growth rate in the share of homes purchased by buyer type z relative to the base period; N_z is the number of pre-period buyers of type z; and p_{jz} is the estimated probability that a type-z buyer purchases in neighborhood j in the pre-period (2012-2016). The first-stage regression takes the form

$$P_{jt} = \gamma_1 E_t[r_j|D_j] + X_{jt}\gamma_2 + \gamma_3 W_{jt} + \varphi_j + u_{jt}$$

The intuition behind the instrument is straightforward. Neighborhoods that were historically popular with buyer types whose national presence is growing face higher demand. Given inelastic housing supply, this demand pressure translates into higher prices. The variation in W_{jt} is driven by national shifts in buyer-type composition that are plausibly unrelated to idiosyncratic changes in neighborhood amenities within California.

The validity of the instrument rests on a conditional exogeneity assumption:

$$E[W_{jt}\xi_{jt}|X_{jt}, \varphi_j] = 0$$

National changes in buyer-type composition must be uncorrelated with neighborhood-specific amenity changes in California, conditional on observed characteristics and neighborhood fixed effects. Including neighborhood fixed effects in the second-stage regression weakens this assumption by restricting identification to within-neighborhood variation in W_{jt} over time, ruling out concerns about correlation between amenities and the choice propensities p_{zj}. The primary remaining threat to exogeneity is the possibility that national shifts in buyer-type composition correlate with amenity changes over time across neighborhoods in California.

6 Results

We now present and interpret the estimates of the demand model. Following the order estimation procedure, we first present the maximum likelihood estimates of the parameters associated with the individual-specific component of utility, and then the IV estimates of the mean value decomposition.

6.1 Disclosure Policy Estimates

Table 3 presents the maximum-likelihood estimates for the disclosure policy parameters for the probit model developed above. Column 1 shows the estimates from the standard probit, which imposes $\pi = 1$, while Column 2 allows the model to freely estimate π alongside the other parameters. The data strongly reject imposing $\pi = 1$, so our preferred specification is Column 2. Accounting for the ceiling probability shifts the disclosure threshold leftward (towards lower-risk neighborhoods) and halves the estimate of the noise parameter τ. Figure 6 plots the estimated disclosure policy rules over a binned scatterplot of the neighborhood risk scores R_j against the probability of disclosure $\Pr(D_j = 1)$.

Table 3: Disclosure Policy Parameter Estimates

	Standard probit	Ceiling probit
$\bar{r}$ disclosure threshold	-1.14	-1.96
	(0.0647)	(0.0621)
τ probit scale	2.11	1.03
	(0.0844)	(0.0610)
π disclosure ceiling	1	0.791
	–	(0.0166)
Log-likelihood	-1,118	-1,049
LR test vs. standard probit		139
p-value, boundary-corrected		< 0.0001
Neighborhoods	2,263	2,263

Notes. Maximum likelihood estimates of the disclosure rule $\Pr(D_j = 1 \mid R_j) = \pi\,\Phi((\log R_j - \bar{r})/\tau)$, estimated on the cross-section of neighborhoods, where R_j is the continuous wildfire risk index (USFS wildfire hazard potential). In the model buyers apply the same rule to latent risk r_j, which the normalization $\lambda_r = 1$ places in the units of $\log R_j$. $\bar{r}$ is the threshold on that log scale, τ governs how sharply the disclosure probability rises around it, and π is the highest attainable disclosure probability. The standard probit imposes $\pi = 1$. Neighborhoods with a hazard score of zero have $\log R_j$ undefined and are dropped. The restriction $\pi = 1$ lies on the boundary of the parameter space, so the likelihood ratio statistic is not χ^2_1 under the null; its limiting distribution is the mixture $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$.

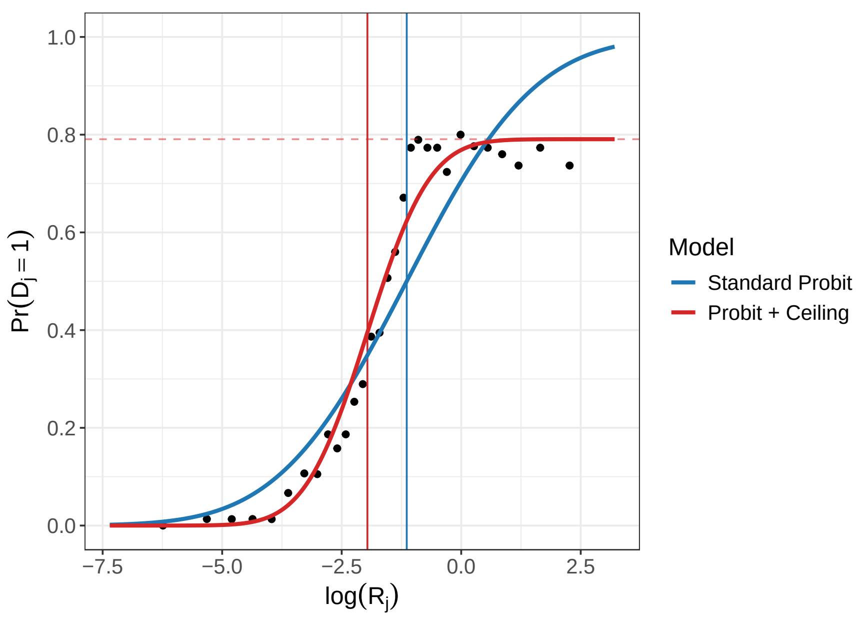


Figure 6: Estimated Disclosure Policy Rules

6.2 Demand Model Estimates

Table 4 presents the estimated parameters from the demand model using the two-stage estimation procedure described in section Section 5.

Panel A presents the estimates of risk belief sensitivity parameters θ_k^R as well as λ_F, the loading of

Table 4: Demand Model Estimates

	Estimate
Panel A: Prior belief formation	
$\bar{\theta}^R$ baseline risk sensitivity	-0.0138
	(0.00182)
θ_1^R type-1 deviation	0.0154
	(0.00284)
θ_2^R type-2 deviation	0.0213
	(0.00381)
θ_3^R type-3 deviation	0.0430
	(0.00484)
λ_f cumulative fire-year, in units of $\log R_j$	1.82
	(0.330)
Panel B: Belief updating	
σ prior standard deviation	13.4
	(0.435)
Panel C: Price sensitivity	
$\bar{\theta}^P$ baseline price sensitivity	-1.47
	(0.168)
θ_1^P type-1 deviation	-0.00556
	(0.00383)
θ_2^P type-2 deviation	0.0117
	(0.00705)
θ_3^P type-3 deviation	-0.00632
	(0.00412)
Implied own-price elasticity, type 0	-1.47
range across types	[-1.48, -1.46]
First-stage F	90.1
Neighborhood-years	15,506

Notes. Point estimates and standard errors from the two-step estimator. First-step standard errors are computed from the inverse observed information; the standard error on σ is from the curvature of the profile likelihood. Standard errors from the second-step IV regression are clustered at the neighborhood level. The prior-belief block (Panel A) is estimated on the pre-disclosure period, where σ does not enter the likelihood; σ (Panel B) is then estimated on the post-disclosure period with the prior block held fixed. $\bar{\theta}^P$ is the baseline price sensitivity absorbed into δ_{jt}, recovered using an IV regression on the pooled 2017–2023 panel of mean utilities. λ_r is normalized to 1 to scale-identify the belief formation block in the pre-policy period, which measures latent risk r_j in the units of the log risk index $\log R_j$; λ_f converts one cumulative fire-year into those units. Price is measured in logs, so $\bar{\theta}^P + \theta_k^P$ is the own-price elasticity up to a factor $(1 - s_{jt}) > 0.99$. The standard error on σ treats the prior-block estimates as known and so understates its uncertainty.

realized fire exposure onto the location of the mean of the prior belief distribution, measured in log-WHP scale units. A positive estimate of the type-specific deviations θ_k^R implies that group k has a lower sensitivity to risk belief than buyers in the baseline category (mortgaged buyers of primary homes). We find a clear cut pattern in the type-specific risk sensitivities. Cash buyers are always less risk sensitive than mortgaged buyers, and buyers of primary homes are always more risk sensitive than buyers of secondary homes. The fact that $\theta_k^R > 0\ \forall k$ suggests that mortgaged buyers of primary homes are the most risk sensitive group. The next most risk-sensitive group is mortgaged buyers of secondary homes. This pattern suggests that financing status (mortgage vs. cash) is a stronger determinant of buyer risk attitudes than primary vs. non-primary home usage.

To benchmark the scale of the risk sensitivity parameters θ_k^R, we convert the estimates into the willingness of different buyer types to pay to avoid a given unit of perceived wildfire risk exposure.[8] evaluating at the average home value in the sample of about \$595,000, types 1, 2, and 3 are respectively willing to pay about $\$11,000$, $\$16,000$ and $\$32,000$ *less* to avoid an additional year of wildfire exposure compared to the baseline type, mortgaged buyers of primary homes. These differential WTP figures are robust to the absolute level of $\bar{\theta}^R$, which is estimation is sensitive to the source of within-neighborhood variation in beliefs about wildfire risk that is used to identify the parameter in the IV estimation.

The estimate $\lambda_F = 1.82$ suggests one year of additional wildfire exposure in a neighborhood increase the mean of prior beliefs about neighborhood risk by the 1.82 log-WHP units, which is about 0.81 standard deviations of the risk score. A variance decomposition of prior beliefs across neighborhoods finds that at the start of the sample in 2017, the risk score contributes 59.7% of the variation in the prior compared to 40.3% for the fire exposure component; by 2020 (the end of the pre-policy period) the fire exposure component has overtake then risk score contribution - the risk score accounts for 42.9% of the variation while the fire component accounts for 57.1%. The increase in the fire component over time is somewhat mechanical, as the fire exposure measure is cumulative while the risk score remains static throughout the sample period.

Panel B presents the model estimate of σ_r, the measure of household uncertainty in the prior belief. Together with τ, the noise in the disclosure policy rule, this parameter governs the extent to which households update their beliefs in response to the disclosure signal D_j in the post-policy period. Compared to $\hat{\tau} = 1.03$, $\hat{\sigma_r} = 13.40$ is quite large. This implies that households place a large weight on the signal

[8] The approximate formula for the differential willingness to pay of type k to avoid a log-WHP unit of wildfire risk simplifies to $\theta_k^R/\bar{\theta}^P$, which gives the WTP in terms of the percentage of home value.

and update their beliefs about wildfire risk substantially in response. Figure 7 shows the change in the distribution of model-estimated beliefs across neighborhood disclosure types between 2020 (the past pre-policy year) and 2021 (the first post-policy year).

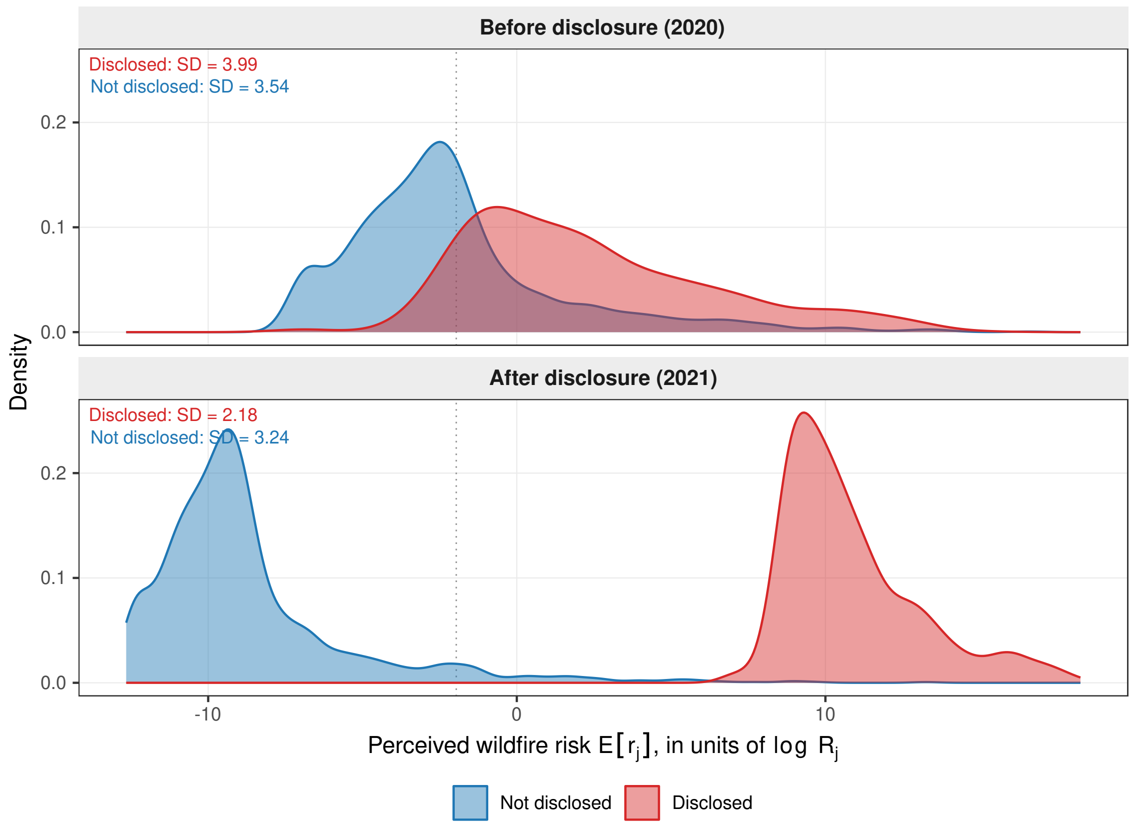


Figure 7: Belief Updating Across Neighborhood Types

Notes: Kernel densities across neighborhoods of perceived wildfire risk, in log-risk units, split by disclosure status. In the 2020 panel beliefs are the prior $\mu_{jt} = \log r_j + \lambda_f F_{jt}$; in the 2021 panel they are the ceiling-aware posterior $\mathbb{E}[r_j \mid D_j, R_j, F_{jt}]$, both evaluated at the model estimates. The dotted line marks the estimated disclosure threshold $\bar{r}$. Adjacent years are used rather than pooled periods because cumulative fire exposure grows over the window, so a pre-post comparison would confound disclosure with fire accumulation.

The disclosure policy has two effects on beliefs. First, it drives a large divergence in beliefs between neighborhoods with $D_j = 1$, where estimated beliefs increase sharply; and neighborhoods with $D_j = 0$, where beliefs fall. Additionally, disclosure reduces the within-neighborhood type dispersion in beliefs; the standard deviation of beliefs falls by 51% in the set of disclosed neighborhoods, and by 16% in non-disclosed neighborhoods.

Panel C of the table reports the estimated price sensitivity parameters θ_k^P. The price sensitivity is about -1.47, which corresponds to the neighborhood-level own price elasticity up to a factor of $1 - s_{jt} > 0.99$ due to the small neighborhood-level market shares. This estimate comes from the second-stage IV decomposition of the mean values δ_{jt}.[9] The IV is well-powered (first-stage $F = 90.1$), and increases the magnitude of the estimated price sensitivity by almost 80% compared to the OLS estimate. This is the expected direction of bias resulting from a correlation between prices P_{jt} and amenities ξ_{jt}. The price sensitivity parameters θ_k^P capture the difference price sensitivity of types $k = 1, 2, 3$ relative to the left-out category, mortgaged buyers of primary homes. At the moment, the level of heterogeneity across types in price sensitivity is small relative to the estimate for the base category.[10]

6.3 Counterfactual Exercise: No Disclosure Policy

To determine the effects of the disclosure policy on household sorting across neighborhoods and equilibrium housing prices, we conduct a counterfactual policy simulation in which the disclosure policy never comes into effect. In the counterfactual exercise, we allow beliefs to evolve with realized fire behavior F_{jt}, but they do not update in repose to the disclosure policy. We hold all other parameters, including the estimated neighborhood amenities $\hat{\xi}_{jt}$ recovered as residuals from the IV regression, at their estimated values. For the baseline policy exercise, we close the model by assuming short run housing supply is inelastic and hold neighborhood market shares s_{jt} fixed at their observed values. We re-solve for market-clearing prices using the BLP contraction mapping to find the vector of prices that rationalizes the observed market shares under the counterfactual set of beliefs. We evaluate the impact of the policy by comparing prices and sorting patterns of household types across neighborhoods under the counterfactual simulation for the post policy years 2021-2023 to those observed in the data.

Figure 8 shows the effect of the disclosure policy on the sorting behavior of type 0 (mortgaged buyers of primary homes) and type 3 (cash buyers of secondary homes) across disclosed and non-disclosed neighborhood with and without the disclosure policy in effect. The largest effects of the disclosure policy are in disclosed ($D_j = 1$) neighborhoods: the disclosure policy leads to a 7.4 percentage point reduction in the share of mortgaged buyers of primary homes, and a 6 percentage point increase in the share of cash

[9] The full set of OLS, first stage, and IV regression estimates are reported in Table B1.

[10] We traced the compression of price sensitivity across types to the our introduction of the outside option of moving out of California. The relatively large outside option shares (ranging from roughly 0.1 to 0.3 across different years seems to have depressed the cross-type price heterogeneity. Two options we are exploring to recover some meaningful degree of heterogeneity in price sensitivity are to explore sensitivity to the assumption about the relative share of out-movers vs. in-movers that are homebuyers, or to add a nesting structure to the outside option where households first decide if they will leave California or stay, and conditional on staying, decide the neighborhood in which they want to live.

buyers of secondary homes. We see the opposite pattern in non-disclosed ($D_j = 0$) neighborhoods, albeit with smaller effect sizes: disclosure increases the share of type 0 buyers in non-disclosed neighborhoods by 2.5 percentage points and decreases the share of type 3 buyers by 2 percentage points.

Figure 8: Changes in Sorting Behavior Under Counterfactual

Notes. Kernel densities across neighborhoods of the share of buyers of each type in 2021, split by disclosure status. Types 0 and 3 bracket the belief weight $\bar{\theta}^R + \theta_b^R$, which is -0.014 for type 0, the only negative value, and $+0.029$ for type 3, the largest. The data series is the observed 2021 equilibrium, with buyer-type shares implied by the estimated model and beliefs given by the ceiling-aware posterior; the counterfactual switches off the disclosure signal so that beliefs revert to the prior mean $\mu_{jt} = \log R_j + \lambda_f F_{jt}$, which continues to evolve with realized fire exposure. Estimated parameters and ξ_{jt} are held fixed. Housing supply is inelastic, so each neighborhood's market share is pinned at its observed level and prices adjust to clear. Filled solid densities are the observed equilibrium and dashed outlines the counterfactual; color indicates disclosure status. Annotations report the mean in the data minus the counterfactual mean; horizontal scales differ across columns.

The re-sorting of buyers across neighborhoods induced by the disclosure policy (which is only an information treatment) changes the incidence of wildfire risk exposure. Table 5 reports changes in exposure to wildfire risk as proxied by neighborhood WHP score and share of buyers in disclosed neighborhood. Mortgaged buyers of primary home reduce their exposure to wildfire risk along both measures, while other buyer types see increases in risk exposure, with the largest effects for cash buyers of secondary homes. This result highlights that information provision alone can alter the incidence of exposure to physical hazards.

The sorting results are driven primarily by the across-type differences in risk sensitivities for the different buyer types, and these effects are unaffected by the level of $\bar{\theta}^R$. In contrast, the general equilibrium price

Table 5: Changes in Wildfire Risk Exposure Across Buyer Types

	Mortgage		Cash		All
	Primary	Non-Primary	Primary	Non-Primary	buyers
Panel A: Mean $\log R_j$					
Data	-2.49	-2.30	-2.23	-1.92	-2.34
Counterfactual	-2.44	-2.31	-2.27	-2.08	-2.34
Difference	-0.0499	0.0104	0.0372	0.156	0
Panel B: Buyers in $D_j = 1$ *neighborhoods (%)*					
Data	14.6	18.8	20.6	28.5	18.2
Counterfactual	17.1	18.4	18.9	21.0	18.2
Difference	-2.54	0.417	1.72	7.55	0
Panel C: Neighborhood burned in purchase year (%)					
Data	8.58	9.91	10.3	12.7	9.64
Counterfactual	8.83	9.85	10.0	11.9	9.64
Difference	-0.255	0.0683	0.206	0.823	0
Share of buyers (%)	55.7	20.1	7.92	16.2	100

Notes. Wildfire risk faced by each buyer type in the observed 2021–2023 equilibrium (Data) and in the counterfactual without the disclosure requirement, in which prior beliefs continue to evolve with realized fire exposure F_{jt}. Difference is data minus counterfactual, in log points in Panel A and percentage points in Panels B and C. Panel C is the share of buyers whose neighborhood was intersected in the year of purchase by a wildfire perimeter of at least one acre, excluding prescribed burns, from the CAL FIRE FRAP perimeter data; these fires do not enter beliefs, since F_{jt} counts fires through year $t-1$. Each neighborhood-year is weighted by the number of buyers of the given type, the product of the market size, the neighborhood's migration-adjusted market share, and the type's share of the neighborhood's buyers, so the figures are pooled over buyers rather than over years. Share of buyers is the type's share of in-state purchases in the data. Housing supply is inelastic, so each neighborhood's market share is held at its observed level and only the composition of its buyers changes; the all-buyers column is therefore identical in the two scenarios. The 2,297 neighborhoods in the simulation include those with a hazard score of zero, which enter $\log R_j$ at $\log(10^{-3})$ as in estimation. Estimated parameters and ξ_{jt} are held fixed.

effects are quite sensitive to the absolute level of risk sensitivity. Table 6 shows the change in prices between the no-disclosure counterfactual and the observed data for disclosed and non-disclosed neighborhoods for a variety of values of $\bar{\theta}^R$, the sensitivity to perceived risk for the baseline group.

The first row, $\bar{\theta}^R = -.0138$, comes from the pooled IV specification where the identification comes from within-neighborhood changes in beliefs caused by realized fire events as well as the disclosure. With this specification, price effects are muted because the non-baseline buyer types all have *positive* willingness to pay for wildfire risk, so prices need to adjust little in order to clear the market. The last row, $\bar{\theta}^R = -.0248$ is estimated using only the within-neighborhood variation caused by realized fire exposure. With this value, most buyer types have negative WTP for fire risk and neighborhood prices have to adjust

Table 6: Counterfactual Price Effects

	$\bar{\theta}^R$	$D_j = 1$	$D_j = 0$	DiD
Joint IV, main table	-0.0138	0.618	0.808	-0.190
Calibrated to a -4% DiD	-0.0178	-1.34	2.66	-4.00
Pre-period only	-0.0248	-4.78	5.93	-10.7

Notes: Effect of the disclosure policy on prices in the counterfactual simulation, under three values of the baseline belief weight $\bar{\theta}^R$. Entries are data minus counterfactual, so a positive value means disclosure raised the price; DiD is the effect in $D = 1$ neighborhoods minus the effect in $D = 0$. Figures are percentages of house value, pooled over 2021–2023 and weighted by observed neighborhood share. The sorting results from the same simulation are invariant to $\bar{\theta}^R$, since it is type-invariant and absorbed by the price adjustments.

substantially in order to induce the less risk sensitive types to sort back into the neighborhood. The middle row, $\bar{\theta}^R = -.0178$, lies between and starts from the estimated diff-in-diff effect of disclosure on home prices from Bass and Fogarty (2025) of -4% and backs out the required level of risk sensitivity in order to match that estimate from the counterfactual simulation.

Figure 9 shows the spatial distribution of the price effects across the different values of $\bar{\theta}^R$ from Table 6. As the absolute value of $\bar{\theta}^R$ increases, the size of the price adjustments needed to clear the housing market also increases, but the spatial pattern is similar across all three scenarios. Prices increase the most in areas with low wildfire risk: the Central Valley, desert areas of southeast California, and core urban areas of Los Angeles, San Francisco, and San Diego. In contrast, prices fall sharply in the Sierra Mountains,the North Coast region, and mountainous areas on the urban fringes of the state's major metro areas.

7 Discussion

This paper studies how climate risk information shapes housing market outcomes through its effects on household beliefs and residential sorting. We develop a structural model of neighborhood choice in which households beliefs about neighborhood wildfire risk update in response to a risk disclosure policy, generating substantial equilibrium adjustments in both sorting patters and prices. Disclosure induces systematic sorting along dimensions of price sensitivity and climate risk aversion, reallocating housing in high risk neighborhoods toward households that are less risk averse and more responsive to price. Financing constraints play a central role in shaping these dynamics, as insurance and lending requirements limit the ability of mortgage-financed buyers to purchase in high risk areas and shift demand toward cash

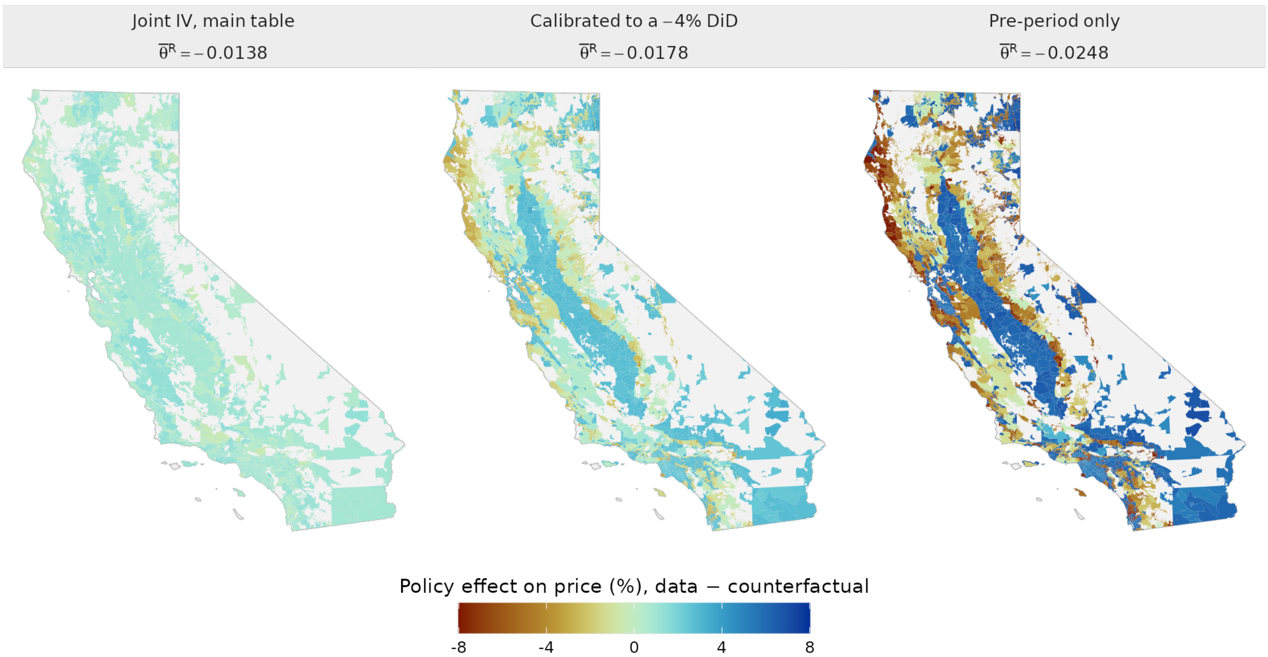


Figure 9: Spatial Distribution of Counterfactual Price Effects

Notes. Each map shows the effect of the disclosure policy on neighborhood prices, $100 \times (\log P_{jt} - \log P_{jt}^{CF})$, the log price index in the data minus its counterfactual value, averaged over 2021–2023 within each of the 2,297 neighborhoods in the simulation. Values are approximately percentages of house value; a positive value means disclosure raised the price. The counterfactual removes the disclosure requirement in 2021–2023 while prior beliefs continue to evolve with realized fire exposure F_{jt}. Estimated parameters and unobserved neighborhood quality are held fixed, housing supply is inelastic so each neighborhood's market share is held at its observed level, and prices adjust to clear the market. The three panels differ only in the baseline risk sensitivity $\bar{\theta}^R$: the joint IV estimate from the main table, the value that reproduces a 4 percent difference-in-differences decline in prices between disclosed and non-disclosed neighborhoods (weighted by market share and pooled over 2021–2023), and the estimate from the pre-disclosure period alone. $\bar{\theta}^R$ is common to all buyer types, so it leaves the simulated buyer composition unchanged and moves only prices. The color scale is symmetric around zero and shared across panels; values beyond ± 8 percent, 1.6 percent of neighborhood-panel values, are shown at the endpoint color. Light gray areas contain no neighborhood in the simulation.

buyers, who become the marginal participants in these markets. Taken together, these results highlight a broader implication for climate risk policy: information interventions do not reduce aggregate exposure to environmental risk, but instead reallocate that risk across households through equilibrium adjustments.

7.1 Sorting Responses to Wildfire Risk Information

Our results provide clear evidence that wildfire risk disclosure induces systematic re-sorting across neighborhoods, driven by heterogeneity in households' preferences over prices and risk. The structural estimates align closely with the reduced-form findings in Bass and Fogarty (2025), which document declines in housing prices in high-risk areas alongside increases in cash purchases and reductions in primary residence purchases. The model rationalizes these patterns as equilibrium responses: when disclosure shifts beliefs about wildfire risk, more risk-averse households, particularly mortgage-financed buyers purchasing primary residences, shift away from high risk areas, while more price-sensitive and less risk averse households move in.

These patterns are consistent with a broader literature on residential sorting under environmental risk. In particular, our findings parallel Bakkensen and Ma (2020), who document that households sort into flood-prone areas along dimensions of price sensitivity and risk tolerance. In both settings, higher risk locations are disproportionately occupied by households that are more responsive to price and less sensitive to environmental risk. However, the dimensions along which sorting occurs differ across contexts. While Bakkensen and Ma (2020) emphasize sorting by income and race, our results highlight the importance of financing status and intended property use. This difference likely reflects both institutional features of the housing market and data limitations: their use of mortgage data restricts their attention to mortgage financed transactions, whereas our data capture the full set of buyers, including cash purchasers.

These differences imply distinct distributional consequences. In our setting, the marginal buyers moving into high risk areas following disclosure are disproportionately cash buyers and purchasers of non primary residences, groups that tend to be whiter, wealthier, and older. This contrasts with settings where lower-income households sort into risk due for affordability reasons. Taken together, these results suggest that while climate risk information consistently induces sorting along risk and price preference margins, the incidence of risk across households depends critically on the institutional environment and the constraints shaping who can respond to price changes. More broadly, our findings highlight that information policies can produce potentially advantageous re-sorting in housing markets, reallocating risk toward the set of households that are both willing and able to bear it.

7.2 Financing Constraints and the Role of Cash Buyers

Our results also highlight the importance of financing status as a key, yet underexplored, dimension of heterogeneity in housing market responses to climate risk. Much of the existing literature focuses on mortgage-financed transactions, often due to data availability, implicitly treating these buyers as representative of the broader market. However, our analysis shows that while cash buyers are a minority of transactions (25% in our sample), they play an important role in shaping housing markets in high-risk areas.

Cash buyers differ from mortgage-financed buyers along several economically meaningful dimensions. Because mortgage lending typically requires proof of homeowners insurance, mortgage-financed buyers face binding constraints when insurance becomes costly or unavailable in high-risk areas. In contrast, cash buyers are less directly constrained by insurance requirements and may be more willing to accept wildfire risk in exchange for lower housing prices. Consistent with this mechanism, we find that cash buyers are more price sensitive and less risk averse than mortgage-financed buyers, making them the primary entrants into high-risk neighborhoods following disclosure.

Although these buyers account for a minority of transactions, they are often the marginal participants in high-risk markets and therefore play a central role in determining equilibrium prices and neighborhood composition. Analyses that exclude non-mortgage transactions thus risk overlooking a key channel through which climate risk is reallocated across households. More generally, our findings suggest that incorporating financing into models of housing demand is essential for understanding how climate risk and related policies shape both market outcomes and the distribution of exposure.

8 Conclusion

Climate change is reshaping the distribution of risk across households, but the allocation of that risk depends not only on underlying hazards, but on what households know when making location decisions. This paper studies how information about wildfire risk affects housing market outcomes by developing and estimating a structural model of residential sorting in which households beliefs about underlying wildfire risk update in response to wildfire events and a risk disclosure policy. Exploiting the introduction of California's wildfire risk disclosure policy, we estimate that beliefs about risk in disclosed and non-disclosed neighborhoods diverge substantially, changing both prices and the composition of buyers.

Disclosure induces systematic re-sorting across wildfire risk: more risk-averse households move away from high-risk areas, while buyers who are more price-sensitive and less risk-sensitive move in. Financing constraints play a central role in determining who these entrants are. Because mortgage requirements tie home purchases to the availability of insurance, they limit some households' ability to buy in high-risk areas and shift demand toward cash buyers, who become the marginal participants in these markets. As a result, the incidence of wildfire risk shifts across buyer types: mortgaged buyers and buyers of primary homes face less exposure, while cash buyers and buyers of secondary residences face correspondingly more.

Together, these results carry a broader implication for climate risk policy. Information interventions do not eliminate exposure to environmental risk; they reallocate it across households through equilibrium adjustments in the housing market. In our context, that reallocation shifts exposure toward relatively advantaged households, those less constrained by financing and insurance requirements, who may be better positioned to bear it.

More broadly, our findings highlight the importance of incorporating information frictions and financial constraints into models of housing demand when evaluating climate risk policies. As governments increasingly rely on disclosure, insurance reform, and other information-based interventions to address climate risk, understanding how these policies interact with market structure will be essential for predicting their effects on both efficiency and equity.

References

Bakkensen, Laura A and Lint Barrage (2022) "Going underwater? Flood risk belief heterogeneity and coastal home price dynamics," *The Review of Financial Studies*, 35 (8), 3666–3709.

Bakkensen, Laura A. and Lala Ma (2020) "Sorting over Flood Risk and Implications for Policy Reform," *Journal of Environmental Economics and Management*, 104, 102362.

Barahona, Nano, Cristóbal Otero, and Sebastián Otero (2023) "Equilibrium effects of food labeling policies," *Econometrica*, 91 (3), 839–868.

Bartik, Timothy J (1991) *Who benefits from state and local economic development policies?*: WE Upjohn Institute for Employment Research.

Bass, Sarah and Michael Fogarty (2025) "Home Financing Responses to Wildfire Risk Information and Insurance Availability."

Baum-Snow, Nathaniel and Lu Han (2024) "The microgeography of housing supply," *Journal of Political Economy*, 132 (6), 1897–1946.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan (2007) "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *Journal of Political Economy*, 115 (4), 588–638, 10.1086/522381.

Bayer, Patrick and Christopher Timmins (2007) "Estimating Equilibrium Models of Sorting Across Locations," *The Economic Journal*, 117 (518), 353–374, 10.1111/j.1468-0297.2007.02021.x.

Baylis, Patrick and Judson Boomhower (2023) "The Economic Incidence of Wildfire Suppression in the United States," *American Economic Journal: Applied Economics*, 15 (1), 442–473.

——— (2026) "Mandated versus Voluntary Adaptation to Natural Disasters: The Case of US Wildfires," *Journal of Political Economy*, 134 (3), 895–948, 10.1086/739331.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995) "Automobile Prices in Market Equilibrium," *Econometrica*, 63 (4), 841–890, 10.2307/2171802.

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

Bin, Okmyung and Craig E Landry (2013) "The Effects of Flood Risk on Housing Prices: Evidence from Recent Hurricane Events," *Journal of Environmental Economics and Management*, 65 (3), 361–376.

Biswas, Siddhartha, Mallick Hossain, and David Zink (2023) "California Wildfires, Property Damage, and Mortgage Repayment," Working Paper 23-05/R, Federal Reserve Bank of Philadelphia.

Boomhower, Judson and Meredith Fowlie (2026) "The Efficiency and Equity Impacts of Risk Classification in Catastrophe Insurance," *Environmental and Energy Policy and the Economy*, 7, 10.1086/738539.

Boomhower, Judson, Meredith Fowlie, Jacob Gellman, and Andrew Plantinga (2024) "How Are Insurance Markets Adapting to Climate Change? Risk Classification and Pricing in the Market for Homeowners Insurance," Working Paper 32625, National Bureau of Economic Research, 10.3386/w32625.

Boomhower, Judson, Meredith Fowlie, and Andrew J. Plantinga (2023) "Wildfire Insurance, Information, and Self-Protection," *AEA Papers and Proceedings*, 113, 310–315.

Collier, Benjamin, Tobias Huber, Johannes Gerd Jaspersen, and Andreas Richter (2023) "Homeowners' Willingness To Insure Flood Risks as Prices Increase," *Available at SSRN 4635177*.

Cook, Cody, Pearl Z Li, and Ariel J Binder (2023) *Where to build affordable housing?: Evaluating the tradeoffs of location*: US Census Bureau, Center for Economic Studies Rochester, NY.

Deryugina, Tatyana and Olga Shurchkov (2016) "The effect of information provision on public consensus about climate change," *PloS one*, 11 (4), e0151469.

Epple, Dennis, Thomas Romer, and Holger Sieg (2001) "Interjurisdictional sorting and majority rule: an empirical analysis," *Econometrica*, 69 (6), 1437–1465.

Epple, Dennis and Holger Sieg (1999) "Estimating equilibrium models of local jurisdictions," *Journal of political economy*, 107 (4), 645–681.

Fairweather, Daryl, Matthew E Kahn, Robert D Metcalfe, and Sebastian Sandoval-Olascoaga (2023) "The Impact of Climate Risk Disclosure on Housing Search and Buying Dynamics: Evidence from a Nationwide Field Experiment with Redfin."

Gibson, Matthew and Jamie T Mullins (2020) "Climate risk and beliefs in New York floodplains," *Journal of the Association of Environmental and Resource Economists*, 7 (6), 1069–1111.

Glaeser, Edward L and Joseph Gyourko (2005) "Urban decline and durable housing," *Journal of Political Economy*, 113 (2), 345–375.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift (2020) "Bartik instruments: What, when, why, and how," *American Economic Review*, 110 (8), 2586–2624.

Hansen, Bruce (2022) *Probability and statistics for economists*: Princeton University Press.

Hino, Miyuki and Marshall Burke (2021) "The effect of information about climate risk on property values," *Proceedings of the National Academy of Sciences*, 118 (17), e2003374118.

Ioannides, Yannis M and Jeffrey E Zabel (2008) "Interactions, neighborhood selection and housing demand," *Journal of urban economics*, 63 (1), 229–252.

Kousky, Carolyn (2010) "Learning from Extreme Events: Risk Perceptions after the Flood," *Land Economics*, 86 (3), 395–422.

Ma, Lala, Margaret Walls, Matthew Wibbenmeyer, and Connor Lennon (2024) "Risk disclosure and home prices: Evidence from California wildfire hazard zones," *Land Economics*, 100 (1), 6–21.

Mulder, Philip (2024) "Mismeasuring risk: The welfare effects of flood risk information," *https://dx.doi.org/10.2139/ssrn.4966795*.

Mulder, Philip and Carolyn Kousky (2023) "Risk Rating without Risk Information," *AEA Papers and Proceedings*, 113, 316–320.

Saiz, Albert (2010) "The geographic determinants of housing supply," *The Quarterly Journal of Economics*, 125 (3), 1253–1296.

Sastry, Parinitha (2026) "Who Bears Flood Risk? Evidence from Mortgage Markets in Florida," *The Review of Financial Studies*, 39 (8), 2317–2357, 10.1093/rfs/hhag030.

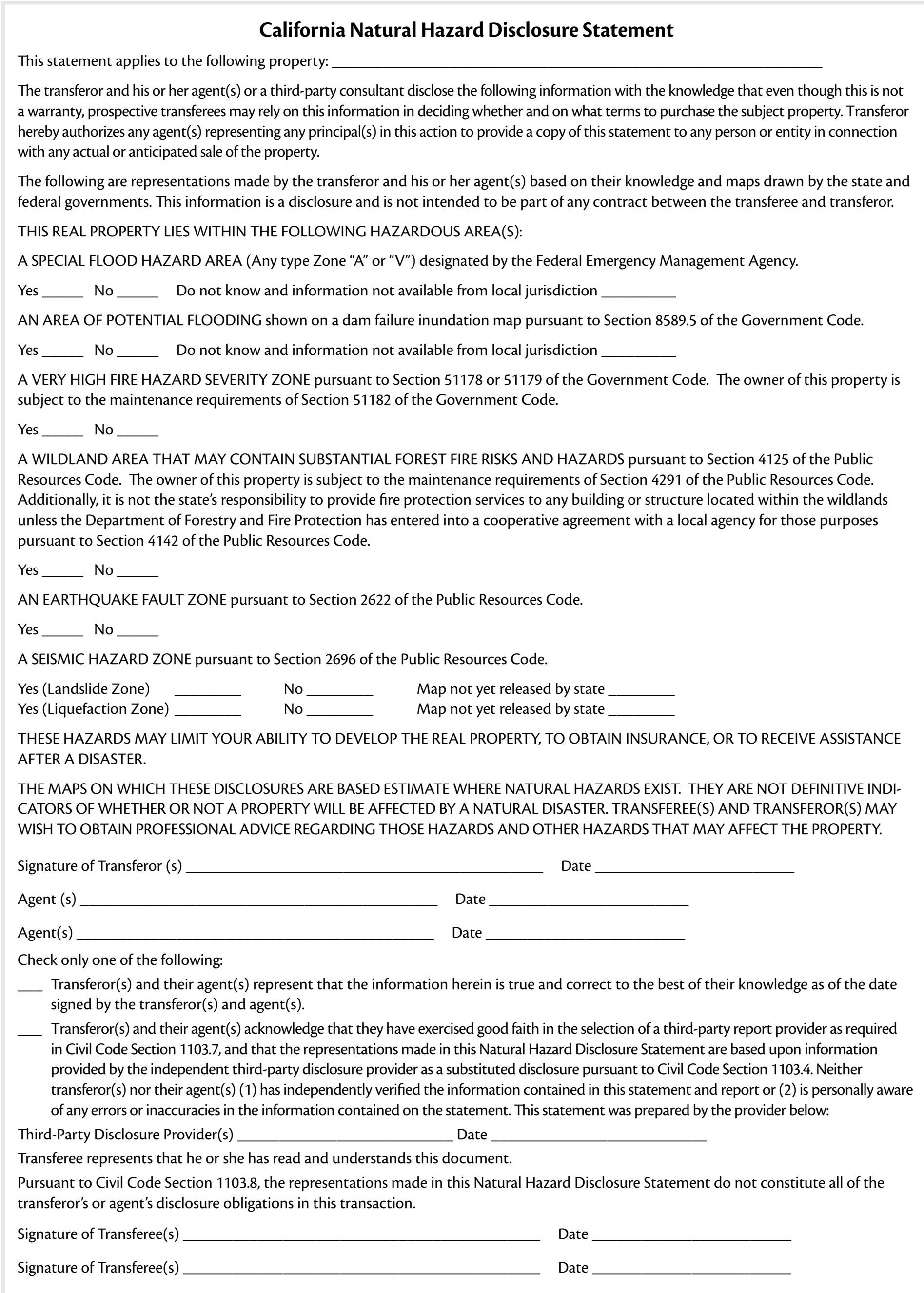
Sieg, Holger, V. Kerry Smith, H. Spencer Banzhaf, and Randy Walsh (2002) "Interjurisdictional Housing Prices in Locational Equilibrium," *Journal of Urban Economics*, 52 (1), 131–153, 10.1016/S0094-1190(02)00007-4.

Weill, Joakim A (2026) "Flood risk mapping and the distributional impacts of climate information," *American Economic Journal: Economic Policy*, 18 (3), 275–305.

Williams, A. Park, John T. Abatzoglou, Alexander Gershunov, Janin Guzman-Morales, Daniel A. Bishop, Jennifer K. Balch, and Dennis P. Lettenmaier (2019) "Observed Impacts of Anthropogenic Climate Change on Wildfire in California," *Earth's Future*, 7 (8), 892–910, 10.1029/2019EF001210.

Wright, Marvin N and Andreas Ziegler (2017) "ranger: A fast implementation of random forests for high dimensional data in C++ and R," *Journal of statistical software*, 77, 1–17.

A Additional Figures



California Natural Hazard Disclosure Statement

This statement applies to the following property: ____________________

The transferor and his or her agent(s) or a third-party consultant disclose the following information with the knowledge that even though this is not a warranty, prospective transferees may rely on this information in deciding whether and on what terms to purchase the subject property. Transferor hereby authorizes any agent(s) representing any principal(s) in this action to provide a copy of this statement to any person or entity in connection with any actual or anticipated sale of the property.

The following are representations made by the transferor and his or her agent(s) based on their knowledge and maps drawn by the state and federal governments. This information is a disclosure and is not intended to be part of any contract between the transferee and transferor.

THIS REAL PROPERTY LIES WITHIN THE FOLLOWING HAZARDOUS AREA(S):

A SPECIAL FLOOD HAZARD AREA (Any type Zone "A" or "V") designated by the Federal Emergency Management Agency.

Yes ______ No ______ Do not know and information not available from local jurisdiction __________

AN AREA OF POTENTIAL FLOODING shown on a dam failure inundation map pursuant to Section 8589.5 of the Government Code.

Yes ______ No ______ Do not know and information not available from local jurisdiction __________

A VERY HIGH FIRE HAZARD SEVERITY ZONE pursuant to Section 51178 or 51179 of the Government Code. The owner of this property is subject to the maintenance requirements of Section 51182 of the Government Code.

Yes ______ No ______

A WILDLAND AREA THAT MAY CONTAIN SUBSTANTIAL FOREST FIRE RISKS AND HAZARDS pursuant to Section 4125 of the Public Resources Code. The owner of this property is subject to the maintenance requirements of Section 4291 of the Public Resources Code. Additionally, it is not the state's responsibility to provide fire protection services to any building or structure located within the wildlands unless the Department of Forestry and Fire Protection has entered into a cooperative agreement with a local agency for those purposes pursuant to Section 4142 of the Public Resources Code.

Yes ______ No ______

AN EARTHQUAKE FAULT ZONE pursuant to Section 2622 of the Public Resources Code.

Yes ______ No ______

A SEISMIC HAZARD ZONE pursuant to Section 2696 of the Public Resources Code.

Yes (Landslide Zone) _________ No _________ Map not yet released by state _________
Yes (Liquefaction Zone) _________ No _________ Map not yet released by state _________

THESE HAZARDS MAY LIMIT YOUR ABILITY TO DEVELOP THE REAL PROPERTY, TO OBTAIN INSURANCE, OR TO RECEIVE ASSISTANCE AFTER A DISASTER.

THE MAPS ON WHICH THESE DISCLOSURES ARE BASED ESTIMATE WHERE NATURAL HAZARDS EXIST. THEY ARE NOT DEFINITIVE INDICATORS OF WHETHER OR NOT A PROPERTY WILL BE AFFECTED BY A NATURAL DISASTER. TRANSFEREE(S) AND TRANSFEROR(S) MAY WISH TO OBTAIN PROFESSIONAL ADVICE REGARDING THOSE HAZARDS AND OTHER HAZARDS THAT MAY AFFECT THE PROPERTY.

Signature of Transferor (s) ____________________ Date ____________

Agent (s) ____________________ Date ____________

Agent(s) ____________________ Date ____________

Check only one of the following:

___ Transferor(s) and their agent(s) represent that the information herein is true and correct to the best of their knowledge as of the date signed by the transferor(s) and agent(s).

___ Transferor(s) and their agent(s) acknowledge that they have exercised good faith in the selection of a third-party report provider as required in Civil Code Section 1103.7, and that the representations made in this Natural Hazard Disclosure Statement are based upon information provided by the independent third-party disclosure provider as a substituted disclosure pursuant to Civil Code Section 1103.4. Neither transferor(s) nor their agent(s) (1) has independently verified the information contained in this statement and report or (2) is personally aware of any errors or inaccuracies in the information contained on the statement. This statement was prepared by the provider below:

Third-Party Disclosure Provider(s) ____________ Date ____________

Transferee represents that he or she has read and understands this document.

Pursuant to Civil Code Section 1103.8, the representations made in this Natural Hazard Disclosure Statement do not constitute all of the transferor's or agent's disclosure obligations in this transaction.

Signature of Transferee(s) ____________________ Date ____________

Signature of Transferee(s) ____________________ Date ____________

Figure A1: Natural Hazard Disclsosure Form

Notes:
This figure displays the updated Natural Hazard Disclosure (NHD) form mandated under California's 2021 wildfire risk disclosure policy. The fields indicating wildfire hazard and wildland areas with substantial forest fire risks are new as of 2021.

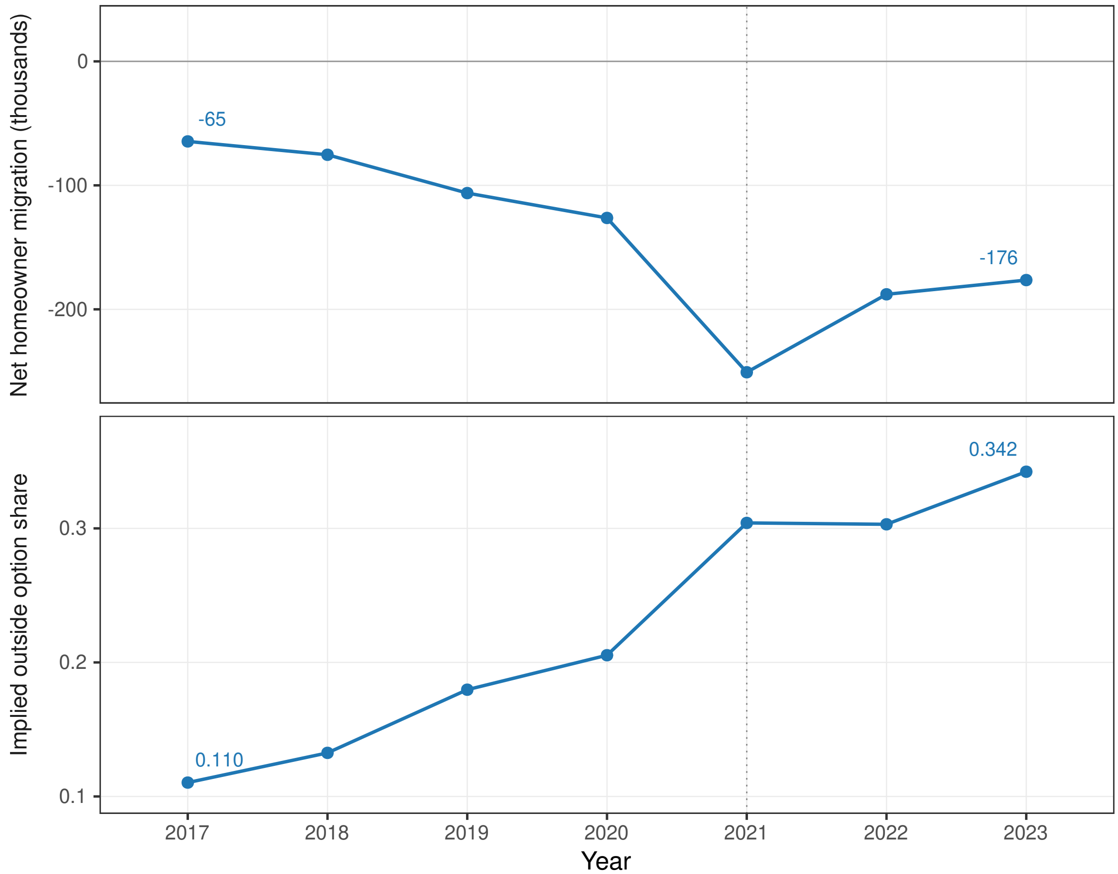


Figure A2: Estimated Out-migration and Implied Outside Option Share

Notes. The upper panel plots net homeowner migration out of California: net domestic migration from the Census Bureau Population Estimates Program (vintage 2020 for 2017–2020, vintage 2025 for 2021–2023), multiplied by the share of out-movers living in owner-occupied housing. The out-mover owner share is computed from ACS 1-year PUMS as the person-weighted share of residents of other states who lived in California one year earlier. The ACS 1-year sample was not released for 2020, so the 2020 share is the mean of 2019 and 2021. Applying the out-mover share to the net flow assumes that in-movers and out-movers have a similar tenure mix. The lower panel plots the implied outside-option share, $s_{0t} = M_t/(N_t + M_t)$, where M_t is the absolute value of net homeowner migration and N_t is the number of in-state residential transactions in the CoreLogic owner-transfer data. s_{0t} is computed as one minus the sum of the migration-adjusted neighborhood market shares, so it equals the outside share in the estimation sample. It rises when out-migration grows or when in-state sales fall. The dotted line marks 2021, the first year of the disclosure requirement.

B Additional Tables

Table B1: Estimation Results: IV Decomposition

	OLS $\hat{\delta}_{jt}$	First stage $\log P_{jt}$	IV $\hat{\delta}_{jt}$
$\bar{\theta}^P$ log price $\log P_{jt}$	-0.823		-1.47
	(0.0448)		(0.168)
$\bar{\theta}^R$ belief b_{jt}	-0.0144	0.00106	-0.0138
	(0.00205)	(0.00100)	(0.00182)
w_{jt} cost shifter, standardized		-0.0273	
		(0.00239)	
Neighborhood fixed effects	Yes	Yes	Yes
First-stage F		90.1	
Within R^2	0.183	0.00596	0.0142
Neighborhood-years	15,506	15,506	15,506

Notes. The second-stage relationship estimated three ways on the pooled 2017–2023 panel of estimated mean utilities. The dependent variable is $\hat{\delta}_{jt}$ in the OLS and IV columns and the log price index $\log P_{jt}$ in the first stage. b_{jt} is the model's belief, $\mathbb{E}[r_j \mid D_j, R_j, F_{jt}]$, entered as an exogenous regressor throughout. Price is instrumented by a standardized shift-share cost shifter. All columns include neighborhood fixed effects; standard errors in parentheses are clustered at the neighborhood level. OLS understates the price response in absolute value, as expected when unobserved neighborhood quality raises both price and mean utility. The belief coefficient is nearly identical under OLS and IV and is unrelated to the instrument in the first stage, so the instrument is not operating through beliefs.

C Hedonic Price Indices

To recover the time-varying neighborhood prices P_{jt}, we estimate the following hedonic regression:

$$\ln p_{kt} = a_{j(k),t} + H_k\Gamma + u_{kt}$$

We estimate the hedonic model on CoreLogic housing transaction data for the state of California from 2017 through Q2 of 2025. We restrict our sample to arms-length transactions of single-family residences, which leaves us with about 3.8 million transactions over the nine-year period. To enrich the hedonic model, we link the transaction data to the CoreLogic property file to obtain rich data on the characteristics of the properties that are sold and the structures on the property.

The goal is to recover estimates of a_{jt}, the neighborhood-time fixed effects. Following Sieg et al. (2002), these fixed effects can be used as the log prices in the sorting model.

Table C1 shows the estimates the hedonic model. In general, coefficients on the continuous covariates all take the expected sign, except for stories, which is negatively associated with price (conditional on square footage, number of bedrooms, and number of bathrooms).

Table C1: Hedonic Regression Results

Dependent Variable:	ln(Sale Amount)
Model:	(1)
Variables	
ln(Square Footage)	0.5421***
	(0.0071)
ln(Acres)	0.0268***
	(0.0028)
ln(Stories)	-0.1140***
	(0.0053)
ln(Bedrooms)	0.0754***
	(0.0062)
ln(Bathrooms)	0.0811***
	(0.0052)
1(Manufactured Home)	-0.3991***
	(0.0225)
1(Pool)	0.0735***
	(0.0050)
Fixed-effects	
ZIP x Year	Yes
Quarter	Yes
Decade Built	Yes
Building Quality	Yes
Garage Type	Yes
Construction Type	Yes
Fit statistics	
Observations	3,811,646
R^2	0.64829
Within R^2	0.17986

Clustered (ZIP x Year) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Our goal in estimating the hedonic model is to obtain the best estimates of the ZIP-year fixed effects. This is distinct from the usual case, where the high-dimensional fixed effects are treated as nuisance parameters and the objects of interest are the coefficients on the "normal" variables in the model. To this end, we make a number of specification choices that are designed to maximize the number of observations that we can use to compute the fixed effects, particularly with respect to our treatment of variables with missing values for some of the characteristics H_k.

For categorical variables (such as building quality or construction type), we simply encode missingness as an additional level of the variable, and the model estimates a distinct intercept for these observations. For continuous covariates (such as building square footage and lot acreage), we first replace missing observation with a common stand-in value, which in practice we implement as the mean of the non-missing values. We also create an auxiliary indicator variable for cases with missing values of that covariate, and include both the original variable with the missing values filled in and the missing value dummy in the model.[11]

Figure C1 plots the estimated ZIP-year fixed effects against the simple within-cell means of the log transaction prices using a binned scatterplot. The blue line overlaid on the data is the best-fit linear regression estimate, while the black line is the identity. The slope of best-fit line is about 0.85, which demonstrates that including the structural characteristics of the housing stock in the hedonic model attenuates the relationship between the mean price and the estimated fixed effects. This suggests that areas with higher housing costs (as measured by the fixed effects a_{jt}) tend to have higher-valued structural features, such as larger lots, more square footage, and more bedrooms and bathrooms. This illustrates the importance of using the hedonic model to estimate the a_{jt} parameters; otherwise in the sorting model we would erroneously attribute to the neighborhood amenities ξ_j variation in prices P_jt that is explained by the characteristics of housing in a a given neighborhood.

Figure C2 compares the (log) distribution of standard errors from the hedonic model (in red) and the standard errors of the ZIP-year cell means (in blue). The standard errors from the hedonic model are about 25% smaller than the standard errors of the means, which illustrates the advantage of including the rich set of housing characteristics for estimating the fixed effects a_{jt}.

[11] Incidentally, both of these procedures for including observations with missing values in the RHS variables tend to *reduce* the R^2 of the hedonic model. This is due to the fact that including extra observations increases the total sum of squares in the dependent variable (log price) without adding much in the way of increasing the explanatory power of the model. If our goal was to maximize the model R^2 or obtain consistent estimates of the Γ parameters, this would be a problem. However, for our particular aim of obtaining the most precise estimates possible of the a_{jt} parameters, the fact that adding these additional observations increases the number of observations in each ZIP-year cell improves precision.

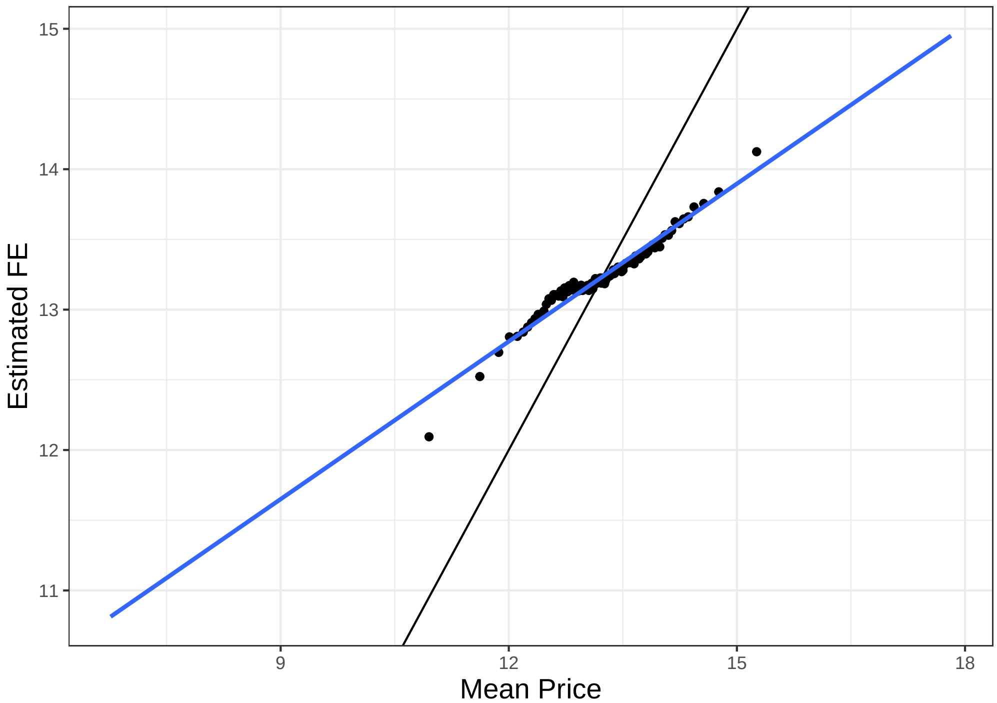


Figure C1: Fixed Effects Estimates

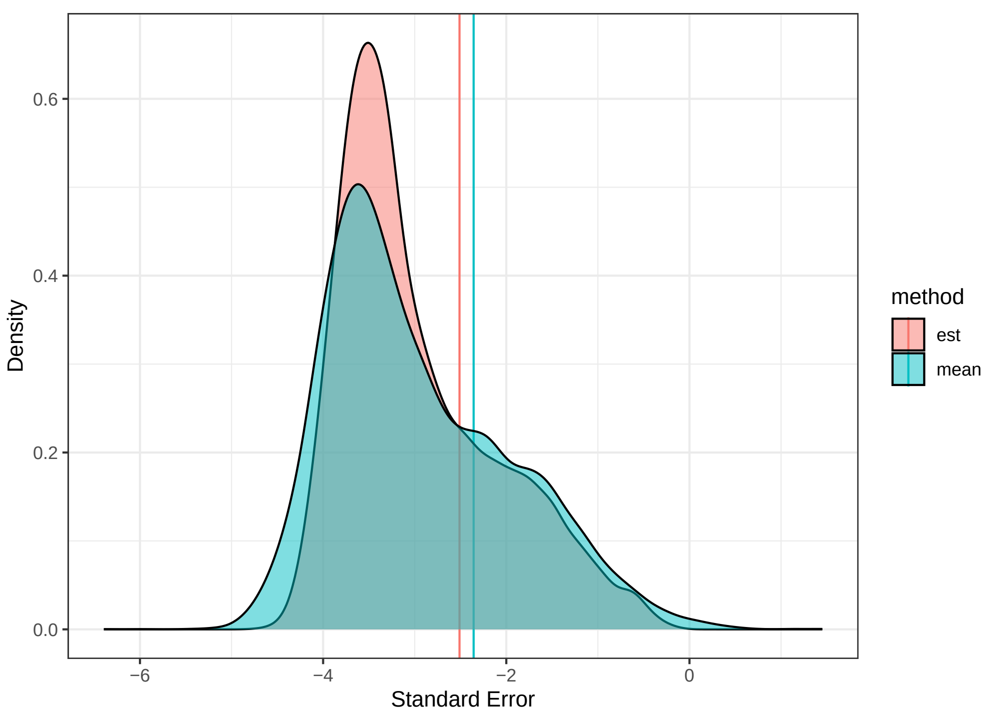


Figure C2: Fixed Effect Standard Error Comparison

C.1 Random Forest

To more flexibly estimate the hedonic model, we generalize the linear model above into the following form:

$$\ln p_{kt} = a_{j(k),t} + f(H_k) + e_{kt}$$

To estimate the model, we proceed in two steps, applying the logic of the FWL theorem to first flexibly estimate the effects of the structural characteristics H_k on prices p_{kt} using a random forest model (Wright and Ziegler, 2017) to estimate the hedonic function $f(\cdot)$.

$$\ln p_{kt} = f(H_k) + u_{kt}$$

Then, having estimated the function $\hat{f}(\cdot)$ via random forest, we construct the first-step residuals:

$$\hat{u}_{kt} = \ln p_{kt} - \hat{f}(H_k)$$

To recover the zip-year fixed effects, we then run the following regression:

$$\hat{u}_{kt} = a_{j(k),t} + e_{kt}$$

Using default settings, the random forest already substantially outperforms the linear model, and after tuning the model hyperparameters, model performance (as assessed by RMSE or R^2) increases further.

C.2 Fixed Effects Decomposition

We also perform a decomposition exercise where we regress the estimated ZIP-year fixed effects on separate zip code and year fixed effects to determine the variance explained by each component (and the residual variance left that is not independently explained by either the time or location FE).

$$a_{jt} = a_j + \tau_t + e_{jt}$$

The model R^2 is about 0.87, which suggests that place and time fixed effects explain most of the variation in the fixed effects from the hedonic model. Of that 0.87, about 0.27 can be attributed to the year fixed effects, while the remaining 0.60 can be attributed to the ZIP fixed effects.

Figure C3 plots the point estimates of the year fixed effects, which can roughly be interpreted as percent growth in average house prices in California compared to a base year of 2017 (the omitted fixed effect category in the model). The estimates suggest that prices grew somewhat modestly across the state from 2017-2020, before growing rapidly between 2020-2022 and leveling off since 2022. These estimates are very consistent with other house price indices, such as the FHFA All-Transactions House Price Index for California.[12]

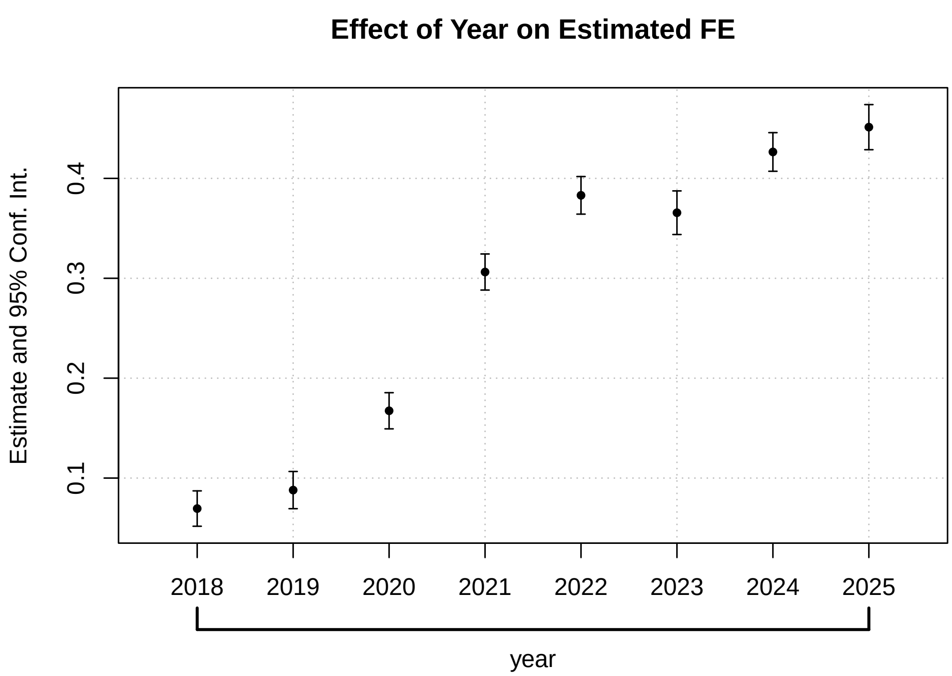


Figure C3: Year Fixed Effects

Figure C4 maps the estimated ZIP fixed effects for the state of California. As expected, prices are higher along the coast and in the major metro areas, and lower inland and in more rural areas. Areas in the map

[12]FRED series `CASTHPI`

without estimated fixed effects correspond to federal landholdings, which tend to be vacant or extremely sparsely populated.

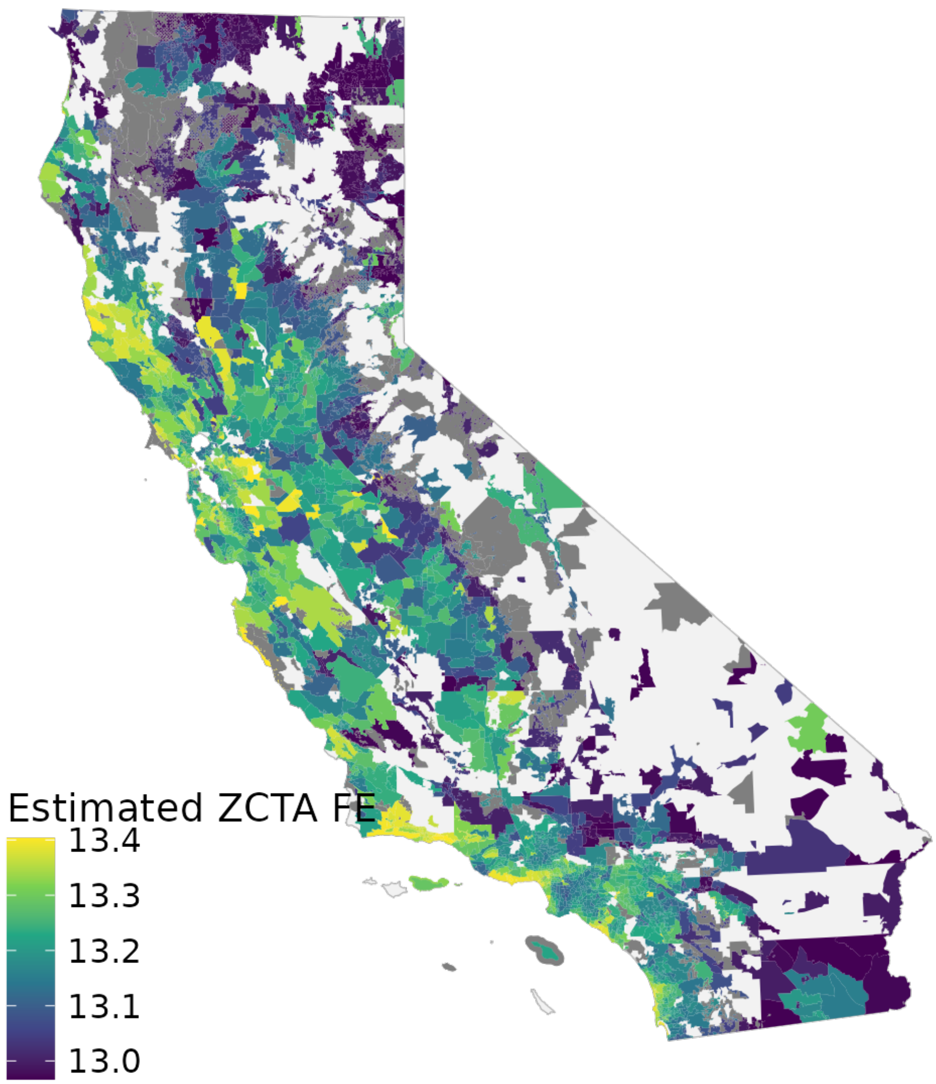


Figure C4: ZIP Fixed Effect Map

D Estimation Details

D.1 Score Vector

To aid in numerical optimization, we derive the score vector — the gradient of the log-likelihood function with respect to the parameter vector $\boldsymbol{\theta}$. For each parameter θ^h, the partial derivative is

$$\frac{\partial \ell(\theta)}{\partial \theta^h} = \sum_t \sum_i \sum_j D_{ijt} \frac{\partial \mu_{ijt}}{\partial \theta^h} - D_{ijt} \frac{\sum_k \exp(\mu_{ikt} + \delta_{kt}) \frac{\partial \mu_{ijt}}{\partial \theta^h}}{1 + \sum_k \exp(\mu_{ikt} + \delta_{kt})}$$

The partial derivatives of μ_{ijt} with respect to each parameter are:

$$\frac{\partial \mu_{ijt}}{\partial \theta^h} = \begin{cases} P_{jt} z_{ik} & \theta^h = \theta_k^P,\ k \in \{1,2,3\} \\ \tilde{R}_{ijt} z_{ik} & \theta^h = \theta_k^R,\ k \in \{1,2,3\} \\ \theta_i^R \frac{\partial \tilde{R}_{ijt}}{\partial \mu_\eta} & \theta^h = \mu_\eta \\ \theta_i^R \frac{\partial \tilde{R}_{ijt}}{\partial \sigma_\eta} & \theta^h = \sigma_\eta \end{cases}$$

D.2 Hessian Matrix

We also construct the Hessian matrix of second partial derivatives, which provides curvature information used by the Newton optimization algorithm and in variance estimation. For diagonal entries ($h = h'$):

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta^h \partial \theta^h} = \sum_t \sum_i \sum_j -D_{ijt} \frac{(\sum_k \exp(\mu_{ikt} + \delta_{kt})(X_{kt}^h Z_{it}^h)^2)(1 + \sum_k \exp(\mu_{ikt} + \delta_{kt})) - (\sum_k \exp(\mu_{ikt} + \delta_{kt}) X_{kt}^h Z_{it}^h)^2}{(1 + \sum_k \exp(\mu_{ikt} + \delta_{kt}))^2}$$

For off-diagonal entries ($h \neq h'$):

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta^h \partial \theta^{h'}} = \sum_t \sum_i \sum_j -D_{ijt} \frac{(\sum_k \exp(\mu_{ikt} + \delta_{kt})(X_{kt}^h Z_{it}^h)(X_{kt}^{h'} Z_{it}^{h'}))(1 + \sum_k \exp(\mu_{ikt} + \delta_{kt})) - (\sum_k \exp(\mu_{ikt} + \delta_{kt}) X_{}}{(1 + \sum_k \exp(\mu_{ikt} + \delta_{kt}))^2}$$

Due to the structure of the buyer-type indicators z_i, a large number of off-diagonal elements of the Hessian are necessarily zero. Specifically, $Z_i^h Z_i^{h'} \neq 0$ only for $h = h'$ (diagonal elements) and the pairs (θ_k^P, θ_k^R) for each type k. The Hessian is therefore sparse and is guaranteed to be symmetric and negative semidefinite.

D.3 Numerical Optimization

We solve the maximization problem using the Newton algorithm. Starting from an initial guess $\boldsymbol{\theta}^0 = \mathbf{0}$, we iterate the update rule:

$$\boldsymbol{\theta}^{h+1} = \boldsymbol{\theta}^h - \alpha H(\boldsymbol{\theta}^h)^{-1} \nabla \ell(\boldsymbol{\theta}^h)$$

The step size $\alpha \in (0, 1]$ is chosen by a backtracking line search to prevent overshooting. We iterate until $\|\nabla \ell(\boldsymbol{\theta})\| < \epsilon$, with $\epsilon = \texttt{1e-8}$. Due to the global concavity of the log-likelihood function, the Newton algorithm converges to the unique global optimum. Results are robust to using derivative-free optimization methods as a check.

D.4 Standard Errors

We follow Hansen (2022, sections 10.15, 10.19) and consider three variance-covariance estimators:

i. $\widehat{V}_1 = \left(-\frac{1}{N} H(\hat{\boldsymbol{\theta}})\right)^{-1}$, the plug-in Hessian estimator.

ii. $\widehat{V}_2 = \left(\frac{1}{N} \mathcal{I}(\hat{\boldsymbol{\theta}})\right)^{-1}$, where $\mathcal{I}(\boldsymbol{\theta})$ is the Fisher information matrix.

iii. $\widehat{V}_3 = \left(-\frac{1}{N} H(\hat{\boldsymbol{\theta}})\right)^{-1} \left(\frac{1}{N} \mathcal{I}(\hat{\boldsymbol{\theta}})\right) \left(-\frac{1}{N} H(\hat{\boldsymbol{\theta}})\right)^{-1}$, the robust sandwich estimator.

Our preferred estimator is $\widehat{V}_3$, which is robust to model misspecification. Standard errors are obtained as $\text{se}(\hat{\theta}) = \sqrt{\text{diag}(\widehat{V}_3)/N}$.