universität wien

MASTERARBEIT | MASTER'S THESIS

Titel | Title

A County-Level Analysis of the Effect of Wildfire Activity on Home Values in California

verfasst von | submitted by

Stephan Wenninger BA

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 913
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Applied Economics
Betreut von | Supervisor:	Ass.-Prof. Dr. sc. ETH Alexandra Brausmann

Abstract (German)

Diese Masterarbeit untersucht den Einfluss von Waldbränden auf die Immobilienpreise in kalifornischen Countys im Zeitraum von 2011 bis 2023. Auf Basis eines Paneldatensatzes für 56 Countys wird die Waldbrandexposition durch den jährlich verbrannten Flächenanteil operationalisiert, während die Immobilienpreise anhand des Zillow-Home-Value-Index für verschiedene Marktsegmente gemessen werden. Die empirische Strategie kombiniert Fixed-Effects-Regressionen mit zeitlich verzögerten Waldbrandvariablen sowie einem quasi-experimentellen Difference-in-Differences-Ansatz, um räumliche und zeitliche Störfaktoren zu kontrollieren. Die Ergebnisse zeigen einen statistisch signifikanten, verzögert eintretenden negativen Effekt auf die Immobilienpreise in allen Marktsegmenten, wobei die stärksten Rückgänge zwei bis drei Jahre nach einem Waldbrand beobachtet werden. Robustness Checks bestätigen, dass dieser Effekt insbesondere von den am stärksten betroffenen Countys ausgeht. Während kurzfristige Angebotsschocks die Preise zunächst stabilisieren können, wird langfristig der Verlust an Standortattraktivität und der damit einhergehende Nachfragerückgang dominant.

Abstract (English)

This thesis investigates the effect of wildfire activity on county-level home values in California between 2011 and 2023. Using a panel dataset covering 56 counties, the analysis operationalizes wildfire exposure as the percentage of county area burned annually and home values via the Zillow Home Value Index (ZHVI) across different market tiers. The empirical strategy combines fixed effects regressions with lagged wildfire exposure and a quasi-experimental difference-in-differences approach to account for spatial and temporal confounding factors. Results show a statistically significant, delayed negative effect of wildfire activity on home values across all market tiers, with the strongest declines observed two to three years after exposure. Robustness checks confirm that this effect is primarily driven by the most severely impacted counties. While short-term supply shocks may initially buffer price declines, the longer-term erosion of amenities and increased perceived risk appear to dominate, leading to falling demand.

Acknowledgement

I want to sincerely thank Ass. Prof. Dr. sc. ETH Alexandra Brausmann for her inputs to this thesis, which were of invaluable help in the writing process.

Also, I would like to express my deep gratitude to my family, who supported me throughout my studies and made this work possible.

Contents

List of Figures

List of Tables

1 Introduction

The US state of California is especially prone to wildfires due to its topography and its dry, windy and warm climate. Climate change has been intensifying the danger of this natural hazard notably over the last decades, largely due to increasing hydroclimate volatility, leading to grasses and shrubs becoming more vulnerable to burning (Swain et al. 2025, McGrath 2025). Of the 20 largest wildfires in California's modern history, 19 have occurred since 2002 — and 13 of those since 2016 (California Department of Forestry & Fire Protection 2025). Research estimating the effect of wildfire activity on home values, with its implications for the wealth of homeowners and the trajectory of the Californian housing crisis (Greenberg et al. 2024), is therefore also becoming more relevant. This thesis aims to contribute by investigating the impact of wildfire activity on county-level home values in California from 2011 to 2023.

Prior research investigating the effect of wildfire activity on home values predominantly finds a negative effect, which is largely attributed to the disamenity that wildfires constitute. Studies focusing on smaller areas in California have specified models to capture not only the development of this impact over time, but also the varying effects on different segments of the home value distribution through employing quantile regression (Mueller / Loomis 2009, Mueller et al. 2014). This thesis also pays attention to these facets by considering different lags of the percentage of area burned in a county as the independent variable and different indices for the lower, middle and upper third of the home value distribution as the dependent variable.

Although most studies on the effects of wildfires on home values are based on the hedonic pricing method (Nicholls 2019), the methodological setup for conducting the estimation, as well as the interpretation of the measured effect, differ in this thesis. The reason for this is that the unit of analysis in focus is not individual house prices, but county-level home value indices. This has implications for the operationalization of the variables capturing home values, wildfires, and other relevant explanatory factors, as well as for the choice of model specifications, including the consideration of spatial effects. When interpreting the findings, the county-level analysis also requires taking into account market effects constituting the measurements that are of lesser importance in the studies based on more granular data and smaller areas. For this, the thesis builds on the work of Boustan at al. (2020), who provide a theoretical framework of the relevant channels to consider when studying the effects of natural disasters on home values at the county level.

This thesis uses panel data regression analysis based on observations ranging from the year 2011 to 2023 in 56 of the 58 Californian counties. The main specification employs county and time fixed effects, as well as multiple control variables. The difference in differences method is then applied by creating a quasi-experimental environment, following the work of Dong (2024), to better identify the impact of wildfire exposure on home values in a county. The findings show a delayed, statistically significant negative effect of wildfire activity on county-level home values across all tiers of the home value distribution. The negative demand-side effects in the county-level housing markets, caused by the destruction of amenities and capital stock, are presumed to overweigh with a time lag, after being offset by the destruction of housing stock limiting the supply in the short-run. The estimated negative effect is mainly driven by the counties most severely hit by wildfires, as further regression analyses on smaller subsets of the sample reveal.

2 Literature Review

Most of the literature on the effects of wildfire activity on home values can be categorized as case studies employing the hedonic pricing method. The hedonic pricing method conceptualizes goods and services as a collection of attributes. Although only one price is paid, consumers value these attributes separately. In the case of houses, attributes could be structural features, such as square footage, but also neighborhood and environmental attributes, such as the crime rate or the proximity to parks. By regressing the overall price of a house on the individual attributes constituting it, the implicit prices for the explanatory characteristics can then be investigated (Nicholls, 2019).

The idea of hedonic pricing being applied to natural hazards originates from the work of Damianos and Shabman (1976), who studied the effects of floodings on residential land values to give a price to this disamenity. Over the recent decades, as Nicholls (2019) shows in her review, this effort has been extended to various other environmental disturbances, such as wildfires. The most common estimation method that is used is cross-sectional OLS, while some studies also employ spatial lags and panel data methods, as addressed in the methodology section. A more detailed discussion of the operationalization of the main variables of interest can be found in the methodology section as well. A short summary of the main findings of the hedonic pricing literature on wildfire risk is provided below.

Many studies have found a significant, negative effect of wildfire activity on individual sale prices of houses. PricewaterhouseCoopers (2001) measured the impact of the Cerro Grande Fire in Los Alamos, New Mexico, in May 2000, employing a regression with pre- and post-fire sales prices of single-family homes. They found a statistically significant average price decline of between 3 and 11% after the fire occurred. Loomis (2004) similarly analyzed the impact of the forest fires in Buffalo Creek, Colorado, on a nearby town, measuring a negative effect of 15 to 16%. For Los Angeles County, California, Mueller et al. (2009) find a decrease of individual house sale prices affected by two consecutive fires, with the second fire's effect (-22.7%) being of greater magnitude than that of the first (-9.7%). Stetler et al. (2010) measured a negative effect (-13.7%) of a wildfire in Northwest Montana on house sales prices within a 5 km radius. In Colorado, McCoy and Walsh (2018) found a negative effect on house prices in a difference in differences estimation approach.

In his literature review, Dong (2024) also points towards interesting findings outside of the United States. In a cross-sectional hedonic model, Paudel (2021) estimated negative effects of wildfire activity on house prices in Nepal. In Australia, Koo and Liang (2022) estimated a significant downward drive of house prices within designated bushfire-prone areas in the state of Victoria, after the occurrence of a major wildfire. Adachi and Lee (2023) investigated the development of house prices situated near the Sampson Flat Bushfire in Mitcham, South Australia. They measure a significant drop of 4.96%.

In contrast to these estimated negative effects, Hansen and Naughton (2013) find significant positive impacts, measuring the effect of large wildfires on assessed home values on the Kenai Peninsula, Alaska. Donovan et al. (2007) used the wildfire risk ratings of the Colorado Fire Department as an explanatory variable for sale prices. They find a positive correlation of riskier areas with prices before the publication of the risk ratings that turns insignificant after the publication. For a case study of Boulder County, Colorado, Rossi and Byrne (2016) employed the number of fires over the past five years in a 1.75-mile radius as a variable to explain sale prices of houses, without finding a significant effect. Hennighausen and James (2024) find a 13 % increase in house prices in a 25 mile radius of the 2018 Camp Fire in Butte county, California, while this effect is decaying with time.

In general, despite the mentioned outliers, most of the case studies that employ the hedonic pricing method suggest a negative impact of wildfires on individual home values. However, one key difference between the hedonic pricing studies summarized above and this thesis is that this thesis does not focus on the granular level of individual house sales data, but the larger-scale effects on county-level home value indices. The modifiable area unit problem (MAUP) should always be appreciated in a context of different levels of analysis, such as this (Manley 2019). It describes the circumstance of obtaining different measurements of the same effects as a consequence of a different level of aggregation or geographical separation undertaken when specifying the unit of observation. Keeping this in mind is important, since the channels through which wildfire activity influences home values at the county level may differ substantially from the smaller-scale hedonic pricing models. The interpretation of the findings has to be adjusted to the more aggregated unit as well.

Turning to a larger unit of analysis, the literature is not nearly as dense as the hedonic pricing case studies. However, Boustan et al. (2020) provide valuable theoretical reasoning behind the effect of natural disasters on economic activity in general, but also on housing prices specifically, while their unit of analysis is US counties as well. They describe various channels through which a natural disaster may impact housing prices positively or negatively. One channel is the contraction of housing stock and therefore a decrease in supply, which leads to a rise in prices when stable demand is assumed. The short-term upward push due to burned structures could be counteracted by new housing built by developers. However, it is also possible that areas are assigned a higher risk factor and that land use regulations are adjusted by the government, which creates persisting supply shortages and keeps prices higher than in the previous equilibrium.

Another channel that Boustan et al. (2020) describe is the decline of demand for living in the county. The occurrence of a natural disaster may lead to a heightened expectation of such an event taking place again. The disaster also may leave devastation that adversely affects the quality of life in a county, for example through the destruction of natural amenities and infrastructure, or heightened air pollution. What is also relevant in this context, is to what level the occurrence of a natural disaster actually changes or fosters the anticipation of a county being vulnerable to this kind of event, or if this perception has been there before. As a third channel, the reduction of labor productivity through destroyed capital and infrastructure may lead to lower wages, followed by a lower inflow and higher outflow in the county's population. This also might depress house prices until the housing stock eventually declines as a consequence.

Considering the smaller-scale hedonic pricing case studies with this theoretical framework as a backdrop, some channels might be relatively underrepresented when focusing on individual sales data close to the fire, since, for example, the supply-side effects driving up housing prices could become more impactful when widening the scope to the county level than negative demand-side effects. Therefore, the interpretation of the findings needs to be adjusted accordingly.

The study of Boustan et al. (2020) stretches over almost a century, while they use decadal data. They find a negative effect of severe natural disasters on housing prices of 2.5 to 5%. However, they explicitly state that the theoretically predicted short-term upward drive due to supply-side effects is likely neglected because of the 10-year time intervals. Dong (2024), whose approach is important for the specification of the models in this thesis, finds close to no effect when focusing on the effect of wildfire activity on home value indices on a neighbourhood level in California.

Following the identification of the differences in interpretation from the studies based on individual housing data, it is crucial to state the value of the county-level analysis to justify this specification. To measure the effect of wildfire activity not only on individual home values, but at the county level, enables an analysis that is relevant due to the implications for larger segments of the Californian real estate market. In this way, the findings can contribute to the discourse and inform policymaking addressing the effects of wildfire activity on home values in California. Contrasting the effects of wildfires over different time lags is especially relevant. While lasting negative effects on home values would have a negative impact on wealth, possible supply-side induced price spikes in the short-term may imply additional financial distress for people having lost their homes to a fire, looking for new housing (Greenberg et al., 2024), as well as people looking to acquire real estate more generally.

3 Data and methodology

3.1 Data overview and operationalization

Wildfires

The first methodological step undertaken is the operationalization of the phenomena of interest by identifying suitable independent and dependent variables. The first operationalization that will be discussed is that of wildfire activity. For this, it is useful to build on the literature presented above. The small-scale hedonic case studies, such as Loomis (2004), Mueller et al. (2009) and Mueller / Loomis (2014), use dummy variables to indicate whether there have been one or two fires close by at the time an individual house is sold. Donovan (2007) uses a similar binary approach, but analyses wildfire risk ratings published by the Fire Department of Colorado Springs, which is his area of focus. Fire radiative power was used by Paudel (2021) to measure the intensity of wildfire activity. Mueller et al. (2009) and Rossi and Byrne (2016) have employed the number of fires in a defined radius as an independent variable. Stetler et al. (2010) have included the size of the closest fire, as well as an indicator for if burned area can be viewed from a house. Since these hedonic models deal with granular sales data of individual houses, it was also possible for many of the studies mentioned to incorporate the distance of a house from the fire into their model.

Because this thesis is interested in a larger scale of analysis, adapting the variables of some of the mentioned hedonic pricing models is not sensible. Using an indicator dummy variable is of little help, since, considering the whole area of a county, there is a high chance of a severe wildfire having occurred somewhere in a recent time period, making it hard to use this as a measure for a severe impact. Using wildfire risk assessment is also not appropriate for this case study. Although the publication of wildfire risk ratings by the fire department has been shown to have effects on house prices (Donovan 2007), these ratings are mostly used to assess differences in risk on a smaller scale. What is more, the literature on wildfire effects on housing prices has shown more promising results when focusing on the actual occurrence of wildfires, rather than risk indicators. Focusing on a larger scale, the neighbourhood level, Dong's (2024) model comprises the share of the area burned in his model as an independent variable. This allows for a more accurate and meaningful operationalisation of the presence of wildfires than a binary indicator, also on a county level, since the magnitude of damage occurred can be taken into account more accurately.

The California Department of Forestry & Fire Protection (2025) releases data of various California fire parameters via the California Open Data Portal, including shape files modeling the area of the fires that occurred during the period of 2011 to 2023 that this study will focus on. For the purpose of acquiring the area burned within a county, which, as established above, is the metric of interest for this thesis, there had to be done some manipulation, since this variable was not provided in the original data set. To obtain this variable, a spatial intersection was done between the fire data and the US county grid downloaded via the TIGRIS package developed by the American Census Bureau (2024). The geometries of the fires and the county borders were reprojected to the coordinate reference system WGS84 to allow for a clip of the fire polygons to the county grid. This operation was done with the Simple Features package (Pebesma 2018) in R. The resulting variable is square meters burned per year from 2011 to 2023 for all 58 Californian counties. This was further processed to calculate the percentage of county area burned, which is used for the regression analysis. As a sanity check, the individual total burned county areas were added up and compared to the yearly total area burned in all of California, which matched.

Many of the studies referenced in the first paragraph of this section include a variable capturing the time elapsed since a major wildfire event. To distinguish between short-term and medium-term effects of wildfire shocks, this thesis introduces lagged versions of the annual percentage of a county area burned, which are incorporated into the main fixed effects specification. These lags will be discussed in more detail in the modeling section. For consistency with this approach, the wildfire data visualizations presented here begin in 2008, which corresponds to the three-year lag preceding the first year of analysis, 2011.

A comprehensive overview of wildfire activity in California from 2008 to 2011 is provided below. Figure 1 shows a mapping of all wildfires, colored by year. Figure 2 and Figure 3 show a summary of the counties with the highest percentage of their area burned, mapped on the California county grid. The five counties most severely exposed to wildfire in the considered time frame according to this measure are Plumas, Lake, Trinity, Napa and Butte.

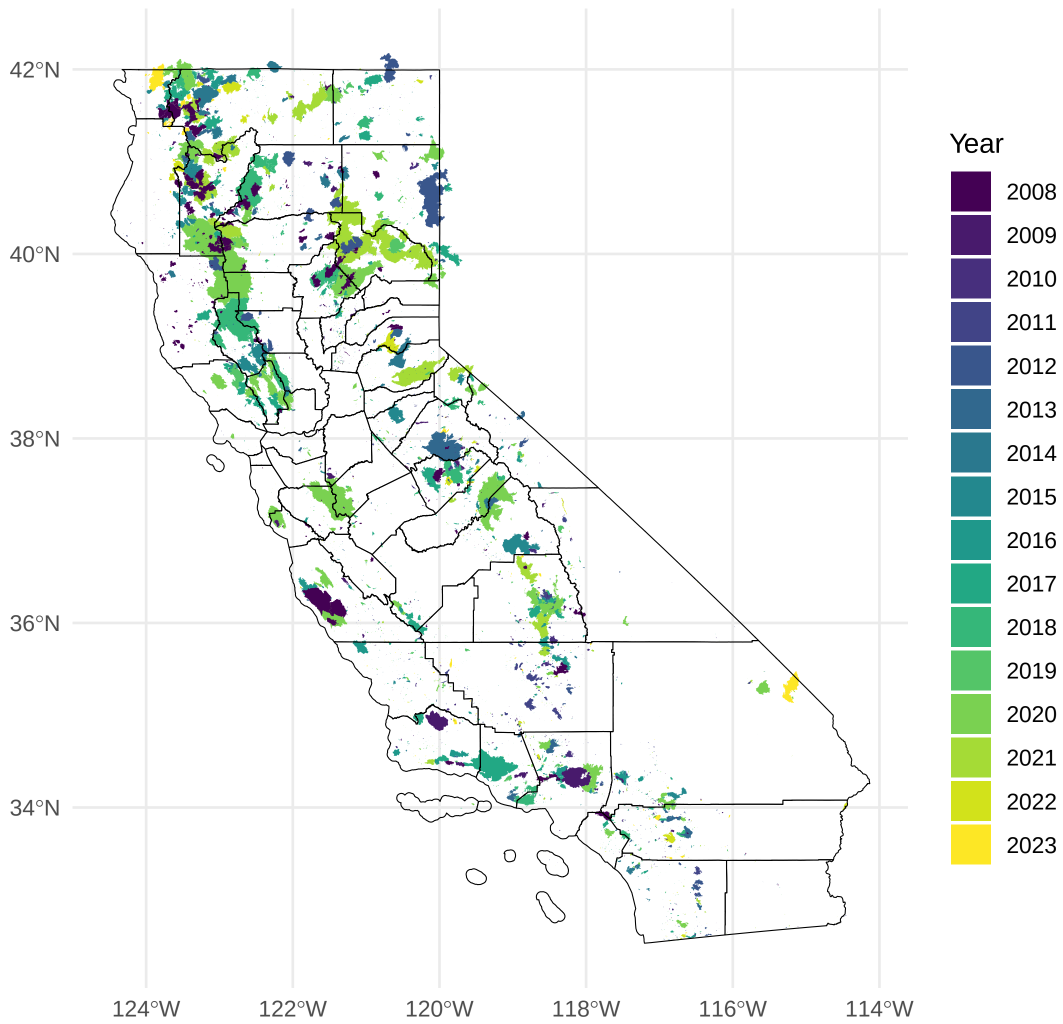

Figure 1: Map of wildfires occurring in California from 2008 to 2023, colored by year.

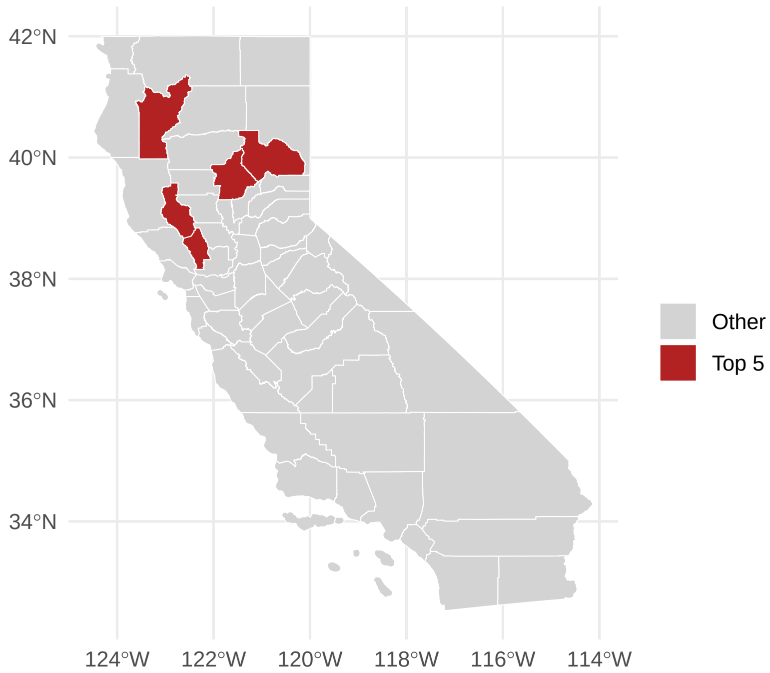

Figure 2: Map of top five counties with most severe wildfire exposure by average yearly percentage of county area burned (2008-2023).

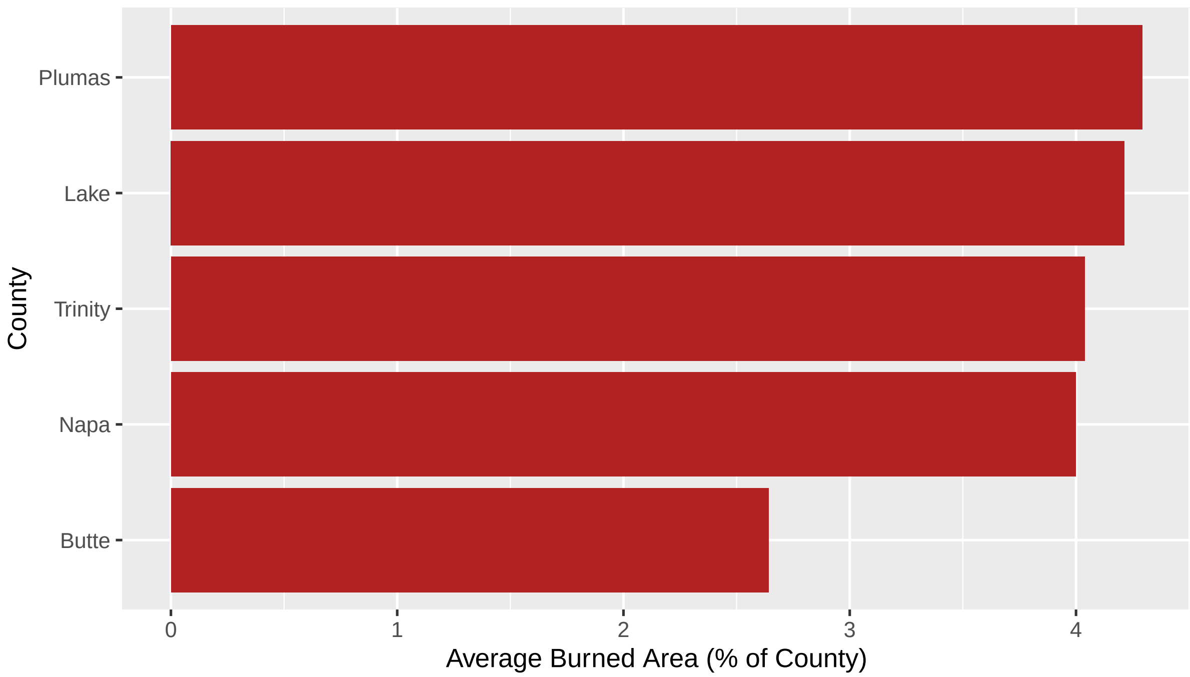

Figure 3: Bar chart of top five counties with most severe wildfire exposure by average yearly percentage of county area burned (2008-2023).

Home values

Now that the independent variable operationalizing wildfire activity has been examined, the dependent variable aiming to identify home values will be discussed. Loomis (2004), Mueller et al. (2009), Mueller / Loomis (2014), Donovan (2007), Rossi, Stetler et al. (2010) and Byrne (2016) have used data on individual house sales for their smaller scale hedonic studies. Dong (2024), on the other hand, focusing on the neighbourhood level, has based his analysis on a home price index created by Bogin, Doerner, and Larson (2019). For the purpose of this thesis, focusing on the county level, the smoothed and seasonally adjusted Zillow Home Value Index (ZHVI) will be used to identify home values. The index has multiple advantages over other operationalization approaches, such as only considering mean or median sale prices, for this aggregate level analysis. Zillow (2025), the leading online marketplace for housing in the United States, estimates not only the value of houses that are on sale, but also houses that have not been listed for long periods of time. Also, the alternative strategy of using the median of housing sales does not take into account the variation in the kind of houses entering the market in a systematic or random way (Olsen, 2023).

Inspired by previous work of Mueller and Loomis (2014), who take a quantile regression approach, this thesis aims to consider a different impact of wildfire activity on different sections of the home value distribution. Since this thesis relies on aggregate data rather than individual sales data, it is not feasible to conduct a quantile regression. However, the ZHVI enables an analysis of a variety of impacts across the value distribution to a certain extent, as explained in the following. The main ZHVI is calculated taking the trimmed mean of the home values by first dropping the bottom and upper third of the distribution. In this way, distortion caused by outliers is circumvented and the measure is hence more adequate in capturing the typical value of a house sold (Olsen 2023). In addition to the measure of this mid-tier section of the housing market, Zillow also releases the index for the low and high tier home values, which each are acquired by taking the trimmed mean of the bottom or upper third of the distribution, respectively (Olsen 2023). Therefore, through using these three different indices as independent variables, wildfire effects on three different tiers of home values in a county can be inferred in the regression analysis.

A map of the different home values in the most recent year of analysis is presented in Figure 4, while summary statistics are provided in Table 1. The map shows a wide range of typical home values in different counties. Over the whole observed period (2011-2023), the average county-level home index for the middle tier is \$528,461 , ranging from \$134,193 to \$1,737,417. Higher home values are found in coastal counties, with especially high values in the San Francisco Bay area.

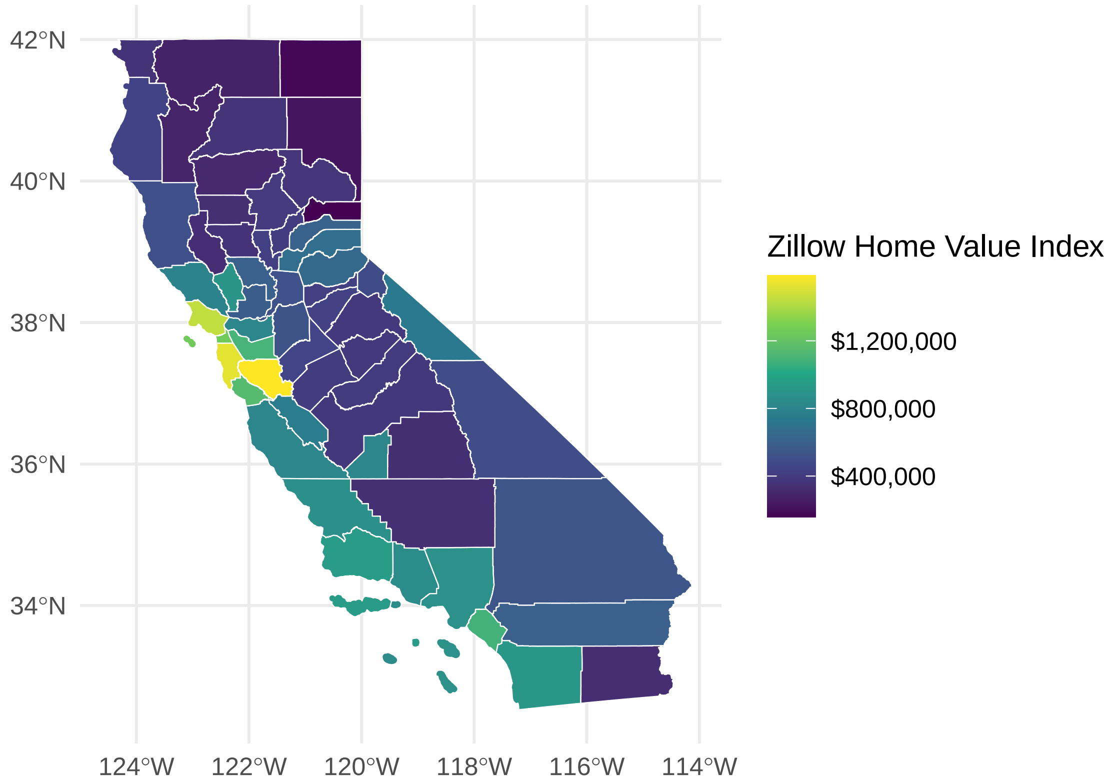

Figure 4: Map of middle tier Zillow Home Value Index in 2023 in different counties.

Insurance Data

Another development of interest when investigating the effects of wildfires on the housing market is that of insurance premiums. However, the limited availability of data described below hinders sound enough statistical inference, which is why the results are only presented in the appendix (Chapter 8.2).

In California, it has been mandatory for insurers with direct written premiums of over $10 million to report the relevant data twice a year to the Insurance Commissioner only since 2018, when Senate Bill 824 was signed to address the lack of historical insurance data (California Department of Insurance 2025). This means that the analysis of the development of insurance premiums in this thesis will be limited to the periods of analysis after 2017. The independent variable of interest is the average insurance premium for each county and year, calculated by dividing the aggregated earned premium by the aggregated earned exposure. The earned premium reflects the total amount of premium revenue insurers have earned by providing coverage during the reporting period, while earned exposure captures the total number of policy-years and units of risk covered. Their ratio hence provides the average premium paid per policy or unit of coverage. The data was downloaded via the website of the California Department of Insurance (2025).

Control variables

Since many confounding factors beyond wildfires are expected to influence the development of home values in California, it is necessary to control for the main explanatory variables affecting housing supply and demand. Given that all models in this thesis account for fixed effects across counties and time periods, the focus lies on variables that may still drive differences in supply or demand across counties while controlling for year-specific shocks. This rationale, along with the selection of specific control variables, is discussed in greater detail in the modeling section.

Again, the studies mentioned in the literature review informed the choice of variables. However, a main challenge was to transfer the operationalization to a bigger level of analysis. For the variables indicating neighborhood qualities, such as the unemployment rate, the shares of educational attainment and median income, this was simply done by acquiring data for the county level. The sizes of the houses, which in the hedonic models could be controlled for by taking the square meters of the individual houses for sale, were taken into account by using the average number of rooms per home in a county as an indicator. To obtain a measure for the age of the structures, the median year of housing units built was subtracted from the last year of the analysis, 2023. The median age of a county's population is also included, since different age groups could be expected to have a different propensity to move, therefore influencing demand elasticity. The net migration rate is introduced to capture possible demand effects in a county, while the housing growth rate should reflect changes in the housing stock, prompting supply side effects. Educational attainment in a county is controlled for by including the high school and academic share, capturing the percentage of the population whose highest qualification is a high school or academic degree, respectively.

Almost all data for the control variables were obtained from the American Community Survey (2025a), which is the leading source for information on housing and population. The data was accessed via the American Census API. While for bigger counties with a population of more than 20,000 people, 1-year estimates for all variables are available, data for smaller counties relies on 5-year estimates to remain reliable, although some noise is expected to be introduced into the analysis due to a slight loss of time accuracy. More information on the difference between the 1-year and 5-year estimates can be found on the American Census Bureau's (2025) website. The migration data was obtained via the website of the Internal Revenue Service (2025), which provides data of migration outflows and inflows of a county. The net migration rate was consequently calculated by dividing the difference of migration inflows and outflows by the overall population at the beginning of a period, the latter being obtained from the American Census Bureau again.

Inflation adjustment

All the prices mentioned above, including home value indices for different tiers, insurance premiums and the median income, were adjusted for inflation. This was done by calculating a yearly inflator, taking the last year of analysis, 2023, as a base year. The price level data for the period of 2010 to 2023 for the western United States region was obtained via the Federal Reserve Economic Data API by the Federal Reserve Bank of St. Louis (2025).

Missing Value Treatment

Since Zillow data for the main specifications is not available from the substantial time period of 2011–2015 for the counties Inyo and Tuolumne, those are excluded in the data used for the fixed effects and difference-in-difference estimations. They are included in the analysis of insurance premiums however, since for this estimation the observed time frame is limited to 2018 to 2023 anyways, because of the incompleteness of insurance data. For the main specification, which focuses on the mid-tier home values as an independent variable, as well as the upper-tier analysis, there is no Zillow Home Value Index data available for Tehama for the year 2011. Even though the panel data set is therefore made slightly unbalanced, Tehama is included in the analysis for the sake of completeness. For the lower-tier, the county Shasta is also excluded due to multiple years of missing data. Additionally to Tehama, Mono is missing data, but only for 2011, and remains in the analysis.

In the panel data of the unemployment variable, 4 missing values have been found, with the counties Lake and Tehama both showing two non-consecutive years with unrecorded data. For the lag of the net migration rate, 7 counties miss values for the first two years of observation. To circumvent excluding these observations, the values were imputed by the forward or backward filling method, respectively. This leads to an estimation on the basis of 56 counties and 12–13 time periods for the main specification fixed effects model of the effect of wildfire activity on middle and high tier home values, amounting to 727 observations, and an estimation based on 55 counties and 12–13 time periods for the lower tier, amounting to 713 observations.

Summary statistics

Summary statistics for all variables discussed above across all counties and time periods are provided in Table 1.

Table 1: Summary statistics (2011-2023) for all variables used in the regression analysis

Variable	$N \times T$	Mean	SD	Min	Max
$ZHVI_{0–33}$ (Lower Tier)	737	351,334.24	221,135.28	61,487.54	1,185,504.61
$ZHVI_{33–66}$ (Mid Tier)	743	528,461.36	318,353.26	134,193.46	1,737,417.18
$ZHVI_{66–100}$ (Upper Tier)	743	866,021.46	558,413.71	225,507.63	2,960,334.46
PctAreaBurned	754	1.21	3.96	0.00	46.73
AverageRooms	754	5.17	0.34	3.90	6.00
MedianHousingAge	754	45.30	7.27	28.00	83.00
UnemploymentRate	754	8.47	3.59	1.90	21.10
RealMedianIncome	754	80,864.20	23,679.13	46,109.79	165,899.79
NetMigrationRate	689	-0.03	0.81	-4.45	4.72
HousingGrowth	754	0.00	0.01	-0.15	0.09
MedianPopulationAge	754	39.66	6.18	29.80	56.00
HighSchoolShare	754	23.58	5.40	9.90	44.60
AcademicShare	754	27.26	11.78	10.40	62.60

3.2 Spatial effects

One of the major concerns about possible biases in the estimation of wildfire effects on home values is the possibility of spatial autocorrelation among the observations. In some of the smaller-scale hedonic property models discussed in the previous chapter, an adequate specification of a spatial model is introduced to reduce bias and improve the efficiency of the estimated parameters. Mueller and Loomis (2010) underscore this necessity in some of their work and numerous analyses (Donovan et al. 2007; Hansen and Naughton 2013; Rossi and Byrne 2016) investigating the effects of wildfires on house prices consider spatial lags in their models. In a DID setup, spatial effects should also be taken into account (Dong 2024).

In general terms, any empirical cross-sectional or panel investigation dealing with observations that share a geographic context should be expected to comprise spatial effects. These spatial effects can broadly be separated into *spatial dependence* and *spatial heterogeneity* (Anselin 2022). A set of observations is spatially dependent if the structure of the correlation between the observations can be inferred from their spatial ordering—in particular, their distance and spatial arrangement. Unlike time-series autocorrelation, spatial autocorrelation is more difficult to model due to mutual dependence among observations, which requires the incorporation of feedback effects. Spatial heterogeneity, on the other hand, is a special form of structural instability that econometricians are familiar with outside of the spatial context. This aspect can be handled with more conventional estimation methods accounting for heterogeneity bias, such as fixed effects. Ignoring spatial dependence or spatial heterogeneity—despite their presence in the true data-generating process—can either lead to biased and inconsistent estimators (in the case of an omitted spatial lag of variables in the model) or to inefficiency with no effect on the bias if spatial effects are present solely within the error term (Anselin 2022).

Following the reasoning discussed above, while spatial heterogeneity can be addressed with panel data estimation techniques such as fixed effects, spatial dependence would require a spatial specification of the model. For the purpose of estimating wildfire effects on home values in different counties, it is therefore sensible to first test for spatial dependence to determine whether such a specification is needed. If the null hypothesis of spatial independence is rejected, one should incorporate spatially lagged variables in the model, as done by Dong (2024), who estimates a spatial lag and a spatial error model in two separate analyses.

Tests for spatial dependence typically involve specification tests comparing a restricted, non-spatial model with one or more defined spatial models. Tests based on maximum likelihood estimation, such as the Lagrange multiplier test, can be used for this purpose. The R package spdep (Bivand and Wong 2018) offers Rao's Score test (equivalent to a Lagrange multiplier test) for spatial autocorrelation and requires a cross-sectional spatial OLS model as input for the unrestricted model. The restricted model in our case is an OLS regression of cross-sectional estimations for different years, that incorporate the same variables as the main specification of this thesis.

The weights matrix used to model the spatial structure of counties is created with the spdep package using a queen contiguity method and is row-standardized (i.e., each row sums to 1). The unrestricted models tested against the restricted one include the following specifications: a spatially lagged dependent variable model (RSlag), a spatial error model (RSerr), and adjusted versions of each that account for both effects (adjRSlag and adjRSerr). The test has been conducted for every year of interest (2015–2023), assuming a consistent weights matrix across panel regressions. The results of the tests for the different years are provided in the appendix (Chapter 8.1).

The null hypothesis of spatial independence cannot be rejected for any of the models consistently, although there are significant rejections in two years for both the RSlag model and the adjRSlag model. This indicates that the specifications including spatially lagged variables or error terms do not consistently perform better than the restricted model. An inclusion of a spatially lagged variable in the main model is therefore not warranted. However, spatial dependence cannot be completely ruled out for our purposes. Thus, it is important to consider spatial dependence in the computation of standard errors to avoid incorrect assessments of statistical significance. This can be achieved using Driscoll-Kraay standard errors (Driscoll and Kraay 1998), as explained in the next section.

The other side of spatial effects—spatial heterogeneity—will be addressed by employing both county fixed effects estimation and a difference-in-differences estimation approach. The details of both estimation techniques in the context of the application are discussed in the modeling section.

4 Modeling

4.1 Fixed effects estimation

The baseline of the variables under consideration can be expected to vary severely among different counties. It is therefore important to avoid a heterogeneity bias in the estimation of the coefficients. This issue is especially delicate in this case study, because, as Dong (2024) notes, the disamenity of wildfires is likely correlated with the presence of natural amenities that positively affect home values. However, since the relevant features, such as distance to the coast, thermal comfort and forest area are relatively stable over time, allowing for county fixed effects of every county is a well-established method to circumvent this heterogeneity.

Simultaneous shocks in a given year to the home value indices of all counties are another relevant concern. Especially the spike after 2020, most likely caused by the pandemic (Dong 2024) is alarming, since 2020 was also a record year for wildfire activity (Figure 5). To prevent an upward bias resulting from this circumstance and other possible disturbances by simultaneous shocks on the whole cross section, year fixed effects are included in the model.

Having taken care of the county and year fixed effects, there are multiple confounding factors expected to persist. To get rid of this remaining endogeneity, control variables are included. When choosing them, the main criterion is their relevance for the changes in supply and demand in the Californian housing market. The hedonic studies focusing on smaller regions of California offer a starting point for the choice of appropriate control variables. Apart from geographical factors, the control variables introduced by Mueller and Loomis (2014) and Mueller et al. (2009) aim to identify size of the individual house, the year the house was built, as well as the educational attainment and income of the area and its unemployment rate. For this case study, the obvious task when controlling for these factors on a county level was to translate the operationalization of control variables to an aggregate level. This process was described in the data overview and operationalization section of chapter 3.

To enhance the interpretational value, all variables, except growth rates and percentages, are transferred to logarithms in order to measure the elasticities between the independent and dependent variables. The main specification is written below (1).

$$
\begin{aligned}
\log(\mathrm{ZHVI}_{it}) = {} & \alpha_i + \lambda_t + \beta_1\,\mathrm{PctAreaBurned}_t + \beta_2\,\mathrm{PctAreaBurned}_{t-1} \\
& + \beta_3\,\mathrm{PctAreaBurned}_{t-2} + \beta_4\,\mathrm{PctAreaBurned}_{t-3} \\
& + \beta_5\,\log(\mathrm{AverageRooms}_{it}) + \beta_6\,\log(\mathrm{MedianHousingAge}_{it}) \\
& + \beta_7\,\mathrm{UnemploymentRate}_{it} + \beta_8\,\log(\mathrm{RealMedianIncome}_{it}) \\
& + \beta_9\,\mathrm{NetMigrationRate}_{t-1} + \beta_{10}\,\mathrm{HousingGrowth}_{t-1} \\
& + \beta_{11}\,\log(\mathrm{MedianPopulationAge}_{it}) \\
& + \beta_{12}\,\mathrm{HighSchoolShare}_{it} + \beta_{13}\,\mathrm{AcademicShare}_{it} + \varepsilon_{it}
\end{aligned} \tag{1}
$$

This model includes county fixed effects (α_i) to account for time-invariant characteristics specific to each county, as well as year fixed effects (λ_t) to control for shocks that are common across all counties in a given year. The coefficient β_1 on $\mathrm{PctAreaBurned}_t$, as well as those on its lags, captures the association between the percentage of county area burned by wildfires and the logarithm of the Zillow Home Value Index, holding all other variables constant. A one percentage point increase in the share of county area burned corresponds to a $\beta \times 100$ percent change in the respective home value index. The net migration rate and the housing growth are lagged to reduce reverse causality effects, which are discussed in the presentation of the results, as well as in the robustness checks section of the thesis.

4.2 Difference in differences estimation

To identify the impact of wildfire activity on home value indices in greater depth, the difference in differences method is employed. This is done by splitting the counties into a treatment and a control group based on whether they were substantially hit by wildfires after a chosen threshold year, or not. Similarly to the quasi-experimental approach that Dong (2024) takes, the observed period is split into two time frames. Both groups are not severely hit by wildfires in the first time frame. The counties in the treatment group suffer major wildfires in the second time frame, while the wildfire impact in the control group remains low.

There are therefore several cutoffs to be decided on. Firstly, the year that splits the observed period into two parts. Based on the overall burned area in California in the years 2011–2023 (Figure 5), as well as an analysis by the California Department of Forestry & Fire Protection (2025b) of the most severe wildfire incidents, a notable rise in wildfire activity can be identified starting from 2017 with a drastic surge in the year 2020. 2017 is therefore a reasonable cut-off year, since it also lies in the middle of the period, leaving

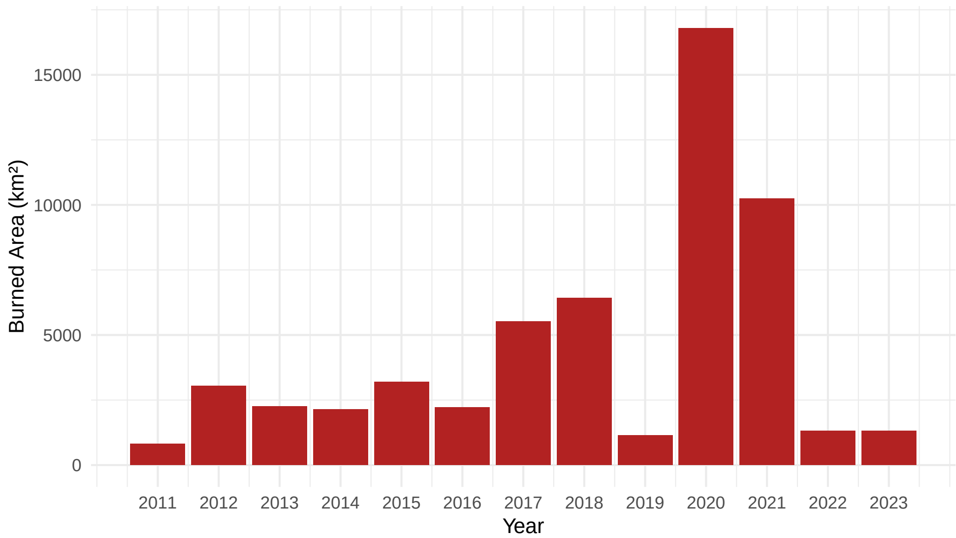

Figure 5: Bar chart of total area burned in square kilometers per year.

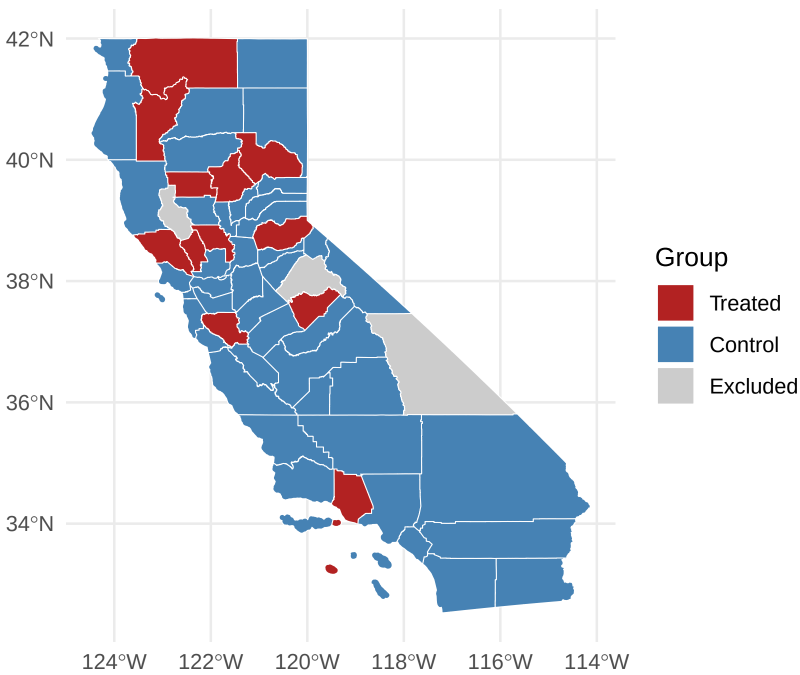

Figure 6: Map of counties in the treatment and control group of the DID analysis.

sufficient observations on each side of the cut-off. A robustness check with 2020 deployed as the threshold year has also been conducted.

The treatment group in the difference-in-differences estimation consists of 12 counties: Butte, El Dorado, Glenn, Mariposa, Napa, Plumas, Santa Clara, Siskiyou, Sonoma, Trinity, Ventura, and Yolo. Two counties—Inyo and Tuolumne—are excluded due to missing Zillow Home Value Index data, as discussed earlier. Lake County is excluded as well, since it experienced substantial wildfire activity prior to the threshold year and therefore does not meet the criteria for either the treatment or control group. The remaining 43 counties form the control group.

Table 2: Balance Table: means and p-values of two-sample t-test (pre-treatment).

Variables	Control	Treated	p-value
ZHVI (mid-tier)	448112.333	452654.462	0.893
Area burned (%)	0.397	0.602	0.294
Real median income	77148.22	78206.293	0.738
Net migration rate (%)	-0.01	-0.064	0.419
Housing growth (%)	0.004	0.004	0.932
Unemployment rate (%)	10.507	10.588	0.864
High school share (%)	23.300	22.678	0.305
Academic share (%)	25.280	28.801	0.007

For both the treatment and the control group, under 3 percent of the county must have been burnt on average per year to be included in the analysis, indicating a relatively lower wildfire activity in the first time frame. In the second time frame, the counties in the treatment group suffered a minimum of 3 percent of the county area burned, representing a major impact of wildfires, while the wildfire exposure in the counties in the control group remained below this 3 percent threshold. The cutoff values were created considering the trade-off between having sufficient counties in both groups, while keeping the severity of wildfire impact clearly identifiable. What was also checked for is the similarity between both groups in the variables of interest. For this, two-sample t-tests have been conducted. As shown in the resulting balance table (Table 2), the only significant difference found is that of the academic share in the two groups. This variable also shows a significant positive impact in the DID regression results. The higher share in the treatment group could lead to a slight upward bias in the DID coefficient. However, since the measured DID effects are significantly negative, at least the direction of the effect remains clearly interpretable, while the magnitude might be slightly understated. What is also crucial for the DID estimation to be valid is that the parallel trend assumption holds before the threshold year, which appears to be the case, as explored in the results section of this thesis.

The DID model can be estimated by expanding the fixed effects regression, creating a DID variable by interacting dummies indicating treatment and post-period observations (2). Also for the DID model the different intercepts are introduced, allowing for year and county fixed effects (3).

$$\mathrm{DID}_{it} = \mathrm{Post}_t \times \mathrm{Treatment}_i \tag{2}$$

$$\begin{aligned}\log(\mathrm{ZHVI}_{it}) = {} & \alpha_i + \lambda_t + \beta_1\,\mathrm{DID}_{it} + \beta_2\,\log(\mathrm{AverageRooms}_{it}) \\ & + \beta_3\,\log(\mathrm{MedianHousingAge}_{it}) + \beta_4\,\mathrm{UnemploymentRate}_{it} \\ & + \beta_7\,\mathrm{HousingGrowth}_{t-1} + \beta_5\,\log(\mathrm{RealMedianIncome}_{it}) \\ & + \beta_6\,\mathrm{NetMigrationRate}_{t-1} + \beta_7\,\mathrm{HousingGrowth}_{t-1} \\ & + \beta_8\,\log(\mathrm{MedianPopulationAge}_{it}) + \beta_9\,\mathrm{HighSchoolShare}_{it} \\ & + \beta_{10}\,\mathrm{AcademicShare}_{it} + \varepsilon_{it}\end{aligned} \tag{3}$$

4.3 Standard error robustness to heteroskedasticity, serial correlation and spatial dependence

For the purpose of finding meaningful results through statistical inference, it is crucial to take care of systematic disturbances in the standard errors, in order not to wrongly assess the hypothesis tests in relation to the studied effects. The first thing to investigate is heteroskedasticity in the error terms, endangering the orthogonality condition of the OLS estimator. A studentized Breusch-Pagan test yields small p-values < 0.01, leading to a rejection of the null hypothesis of homoskedasticity. Furthermore, as discussed in the last section, correlation between different parts of the cross-section should not be completely ruled out despite the negative indications of Lagrange Multiplier tests for spatially lagged models.

Methods established by Driscoll and Kraay (1998) are introduced to allow for a corrected calculation of standard errors that are robust to not only heteroskedasticity, but also serial autocorrelation and spatial dependence. Driscoll and Kraay expand on the widely employed procedure of Newey and West (1987), that accounts for autocorrelation in time series data, by also establishing robustness to systematic disturbances on the cross-sectional dimension. This is achieved through averaging the deviations from the moment conditions of orthogonality across all N in each time period, capturing systematic trends of correlation across the observations by capturing non-random developments over time. This procedure allows for a reflection of systematic spatial disturbances in the estimation of the variance-covariance matrix.

For our environmental panel data case study, the use of Driscoll-Kraay standard errors is useful to address vulnerabilities of the statistical inference presented in the results section. Ignoring not only heteroskedasticity, but possible non-random structures of errors across time and space, may lead to wrong assessments of the significance of our findings (Driscoll / Kraay 1998, Hoechle 2007). Hoechle (2007) introduced this estimator to Stata and made it usable for slightly unbalanced data sets, such as the one used our case. It was incorporated into the Linear Models for Panel Data R-package by Croissant and Millo (2007), which is the version that is used for this thesis.

5 Results

5.1 Fixed effects estimation

The main specification (Table 3) of the fixed effects estimation measures a small negative effect of the share of area burned by wildfires in a year on the middle tier home value index in the same year (December value), as well as one year after the shock. Both effects are statistically insignificant. However, the effect grows in magnitude in the second year after heightened wildfire activity and becomes statistically significant at the 5% level. The measured coefficient suggests that a rise of the area burned in a county by one percentage point is expected to lead to a 0.235% decline in mid-tier home values two years later. The negative impact persists in the consecutive year, with its magnitude slightly declining to 0.228%, while this measurement is statistically significant at the 0.01% level.

The impact on lower tier home values is similar, with a non-significant effect of comparable magnitude in the same year. In this tier, however, the negative effect on home value indices becomes significant as early as the first year after an increase in wildfire activity. However, the estimate of the effect almost doubles in the measurement of the two-year-lag and is statistically significant at the 1% level. As in the mid-tier, there is a highly significant ($p < 0.01$) negative effect of similar magnitude (0.246%) per additional percentage point of area burned on home values three years after an increase in wildfire activity.

The upper tier home values experience a growing, negative effect over the 3 years following an increase in wildfire activity, as well. The same-year effect is again very small and statistically insignificant and remains so in the consecutive year after a shock. The two-year lag coefficient shows a negative effect of similar magnitude (0.23%) as in the other tiers, that is significant at the 5% level. The three-year lag yields a slightly stronger effect than in the other tiers, with a one percentage point increase in burned area leading to a decline of 0.294%. This effect is significant at the 0.1% level.

Overall, the fixed effects estimation shows a lagged, negative effect of wildfire activity in all tiers of the home value distribution. The size of the coefficients differs only slightly across tiers. What is important to consider when interpreting these findings, is that the counties most severely hit by wildfires are the main driver of the effects. This was found in a regression based on a smaller set of counties, excluding counties experiencing particular high wildfire exposure, which is discussed in greater detail in the robustness check section.

The model demonstrates moderate explanatory power, with an R^2 of approximately 0.2 across all specifications. This relatively low value is primarily due to the inclusion of county and even more so year fixed effects, which absorb much of the variation and leave

	Mid Tier	Lower Tier	Upper Tier
PctAreaBurned$_t$	-0.00066	-0.00097	-0.00012
	(0.00075)	(0.00072)	(0.00074)
PctAreaBurned$_{t-1}$	-0.00111	-0.00125^{*}	-0.00099
	(0.00070)	(0.00059)	(0.00071)
PctAreaBurned$_{t-2}$	-0.00235^{*}	-0.00233^{**}	-0.00230^{*}
	(0.00100)	(0.00088)	(0.00095)
PctAreaBurned$_{t-3}$	-0.00228^{***}	-0.00246^{***}	-0.00294^{***}
	(0.00063)	(0.00052)	(0.00079)
log(AverageRooms)	0.22371	0.41242^{*}	0.01869
	(0.14273)	(0.17141)	(0.13767)
log(HousingAge)	-0.16743	-0.19871	-0.10430
	(0.10224)	(0.15338)	(0.08811)
UnemploymentRate	-0.01297^{***}	-0.01386^{***}	-0.00969^{***}
	(0.00277)	(0.00244)	(0.00221)
log(RealMedianIncome)	0.31243^{*}	0.31700^{***}	0.25284^{**}
	(0.12169)	(0.08868)	(0.09481)
NetMigrationRate$_{t-1}$	-0.01631	-0.00637	-0.01278
	(0.01056)	(0.00905)	(0.00887)
HousingGrowth$_{t-1}$	0.34798	0.27973	0.42512
	(0.26528)	(0.28157)	(0.24452)
log(MedianAge)	0.34870	0.27118	0.34877
	(0.31298)	(0.30487)	(0.29154)
HighSchoolShare	0.00656^{***}	0.00468^{*}	0.00481^{**}
	(0.00178)	(0.00203)	(0.00157)
AcademicShare	0.00300^{**}	-0.00037	0.00503^{***}
	(0.00112)	(0.00161)	(0.00105)
R-squared	0.20807	0.20962	0.19309
Adj. R-squared	0.11000	0.10874	0.09316
N	56	55	56
T	$12-13$	$12-13$	$12-13$
Number of obs.	727	699	727

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; standard errors in parentheses

Table 3: Regression table on the effects of wildfire activity on different segments of the home value distribution.

less for the independent variables to explain. When year fixed effects are omitted, the R^2 increases substantially. However, doing so risks biasing the results due to unaccounted-for common shocks across years. Since the aim of this study is statistical inference, rather than predictive power, the decision was made to include year fixed effects in the analysis.

The directions of almost all effects stay constant across the tiers of the home value distribution. A higher average amount of rooms of houses within a county correlates with higher home values, while a higher medium age of housing structures with lower home values. The unemployment rate yields the expected negative coefficient, similarly to estimations in hedonic models (Mueller et al. 2009), and is highly statistically significant. The demographic values of income and educational attainment show a highly significant positive correlation throughout the different tiers, with the exception of the lower tier, where no significant effect of the academic share is detected.

The only surprising effect is the positive correlation of housing growth (except for the lower tier) and negative correlation of the net migration rate with home values. This could be explained by reversed causality, i.e., people moving to places with lower house prices and structures being built in less saturated markets. Both rates, however, show no statistical significance. While different lags for this variable have been tried, also those have not shown a significant effect.

In general, one should be careful when interpreting the magnitude of the different effects, given that percentage rates, including growth rates, are not log-transformed, while other variables are.

Interpretation of the effect

Following the theoretical reasoning behind the effects of natural disasters on home values aggregated on a county-level provided by Boustan et al. (2020), the negative effect measured is rooted in demand-side effects on the real estate market prompted by heightened wildfire activity in a county. The destruction of amenities is the main channel to consider in this context. Lower wages due to the destruction of capital shock, which lead to a lower net migration into a county and hence lower housing demand, may also play a role. The lag of the effect could be explained by supply-side shocks (i.e. the destruction of housing stock) temporarily offsetting the negative demand-side effects. This is in line with the elaborations of Boustan et al. (2020), who attribute a shorter time horizon to the positive supply side effects than to the negative supply side effects.

The estimated impact being predominantly driven by the counties with the highest wildfire exposure, as explained in the robustness checks chapter, suggests that the negative demand-side channels are playing a crucial role only at a relatively high threshold of wildfire activity, or that the positive supply-side effects are asymmetrically strong in counties with more moderate wildfire exposure. Spillover effects of higher demand for housing in the counties surrounding those most severely hit by wildfires are thinkable. This could potentially explain the deterioration of the negative effect in the robustness check excluding the counties with the highest wildfire exposure. However, different spatial lag specifications have been rejected in the Lagrange multiplier tests discussed in the methodology section.

5.2 Difference in differences estimation

The difference-in-differences (DID) estimation allows us to investigate the effect of wildfires more rigorously by creating a quasi-experimental setting through defining a treatment group and a control group, as explained in the modeling section. This helps to further assess the robustness of the negative impact of wildfire activity on home values found in the fixed effects estimation. Additionally, DID enables a visual representation of the treatment effect by plotting the time series of average home value indices of the two sets of counties to investigate the changes in the trends.

The regression results of the DID estimation are presented in Table 4. For the middle tier, there is a statistically significant DID effect of -3.002 %, significant at the 1% level. The effect is significant at the 5 % level for the upper tier, with a size of -2.246 %. The most pronounced treatment effect of -5.424 % is observed in the lower tier of the home value distribution, significant at the 0.1 % level. Considering the time series plots of the average log ZHVIs (Figures 7–9), the common trend assumption of the treatment and control group broadly appears to hold over the years preceding the threshold year 2017 in all three tiers. Although in the lower tier a slight convergence of the trends is visible in the first years of the observed time period, the immediate years before the treatment show a common trend. The robustness checks conducted suggest that the treatment effect in the lower tier is stable and most pronounced also when varying the cutoffs in the treatment assignment process.

The DID results support the findings of the fixed effects estimation. Given that treatment was assigned to counties where the average annual share of burned area was below 3 % before 2017 and above 3 % from 2017 onward, counties in the treatment group are affected by wildfires more severely from 2017 to 2023. Overall, wildfire activity in California increased in 2017 and peaked in 2020 (Figure 5). Although wildfire activity declined after 2020, it remained elevated. Considering the continuous elevated exposure of treated counties to wildfires and the negative effect of wildfires on home value indices measured

	Mid Tier	Lower Tier	Upper Tier
DID	-0.03002^{**}	-0.05424^{***}	-0.02246^{*}
	(0.01061)	(0.01225)	(0.01083)
log(AverageRooms)	0.25710	0.42176^{**}	0.03034
	(0.16031)	(0.16128)	(0.16361)
log(HousingAge)	-0.22651^{*}	-0.26459	-0.15632
	(0.09340)	(0.14691)	(0.08289)
UnemploymentRate	-0.01386^{***}	-0.01591^{***}	-0.01022^{***}
	(0.00309)	(0.00324)	(0.00256)
log(RealMedianIncome)	0.34064^{**}	0.37541^{**}	0.28405^{**}
	(0.12864)	(0.11425)	(0.09896)
NetMigrationRate$_{t-1}$	-0.01409	-0.00613	-0.00940
	(0.01044)	(0.00949)	(0.00898)
HousingGrowth$_{t-1}$	0.38395	0.38306	0.44877
	(0.30889)	(0.38379)	(0.28873)
log(MedianAge)	0.20070	0.18368	0.17220
	(0.27492)	(0.26994)	(0.26212)
HighSchoolShare	0.00670^{**}	0.00472^{*}	0.00465^{**}
	(0.00205)	(0.00193)	(0.00167)
AcademicShare	0.00251^{*}	-0.00165	0.00457^{***}
	(0.00114)	(0.00152)	(0.00110)
R-squared	0.22460	0.24966	0.19421
Adj. R-squared	0.13208	0.15948	0.09807
N	55	54	55
T	$12-13$	$12-13$	$12-13$
Number of obs.	714	701	714

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; standard errors in parentheses

Table 4: Regression table of quasi-experimental difference in difference estimation showing the treatment effect of severe wildfire exposure in 2017-2023.

in the fixed effects regression, a negative DID coefficient was to be expected.

Overall, the negative impact of the DID effect on the average home value trend of the counties in the treatment group is clearly visible in Figures 7–9. While the mean home value index of treated counties was higher or approximately equal to that of control counties in 2017 in all specifications, it is lower or nearly equal by the end of the observed period.

The visualizations further underscore the growing negative effect on home values in the years following heightened wildfire activity. The convergence or, respectively, divergence of the trends becomes more apparent over time after the threshold year. Especially when examining the time series for the treatment and control groups in the mid and upper tiers of the ZHVI (Figures 7 and 9), the divergence suggesting a negative wildfire effect only becomes clear after 2020. It is important to note that the interpretation of the spike in wildfire activity in 2020 (visible in Figure 5) being responsible for this shift in trends is likely misleading. Robustness checks considering 2020 as a threshold year reveal lower and less reliable treatment effects than in the specification employing 2017 as the cutoff. The lagged convergence or divergence of the trends is more likely attributable to the time lag of the negative impact estimated in the fixed effects estimation. This would further support the interpretation of negative demand-side effects notably dominating positive supply side shocks only after a certain time.

The DID estimation yields control variable coefficients similar to those of the fixed effects regression, which is expected since both models use the same independent variables and apply fixed effects for counties and time periods. The most influential explanatory variables are again the unemployment rate, which has a statistically significant negative effect across all tiers, the median income and the variables measuring educational attainment, which generally show significant positive effects. The exception remains the academic share, which shows a non-significant negative effect on the lower-tier home value index.

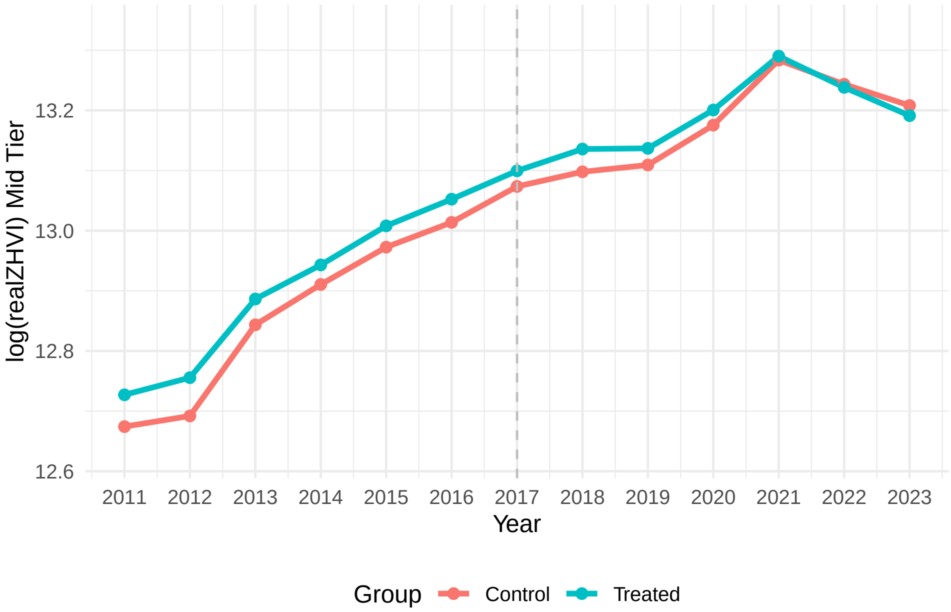

Figure 7: Average Log-transformed Real Zillow Home Value Index (mid tier) by treatment group over time. Dashed line indicates 2017 treatment threshold.

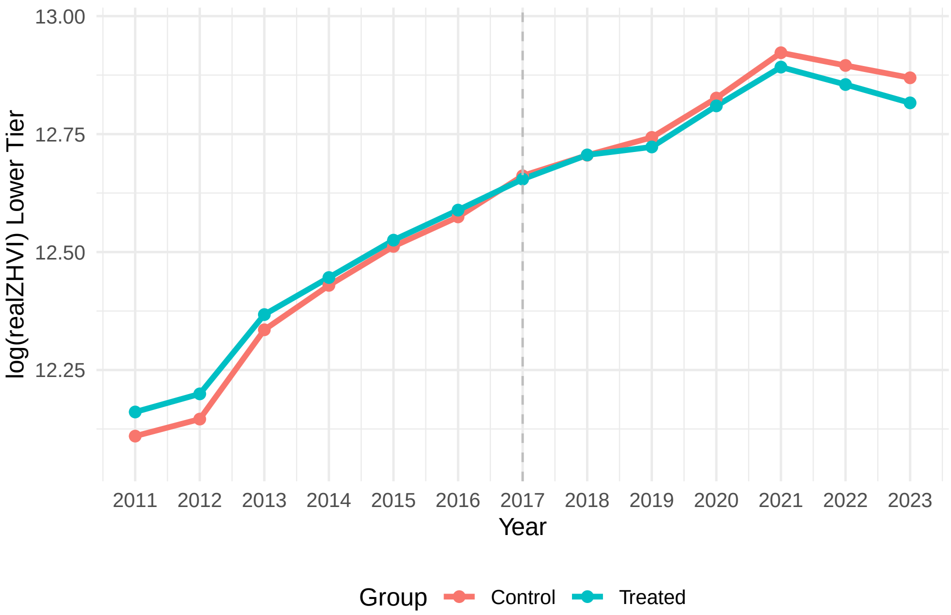

Figure 8: Average Log-transformed Real Zillow Home Value Index (lower tier) by treatment group over time. Dashed line indicates 2017 treatment threshold.

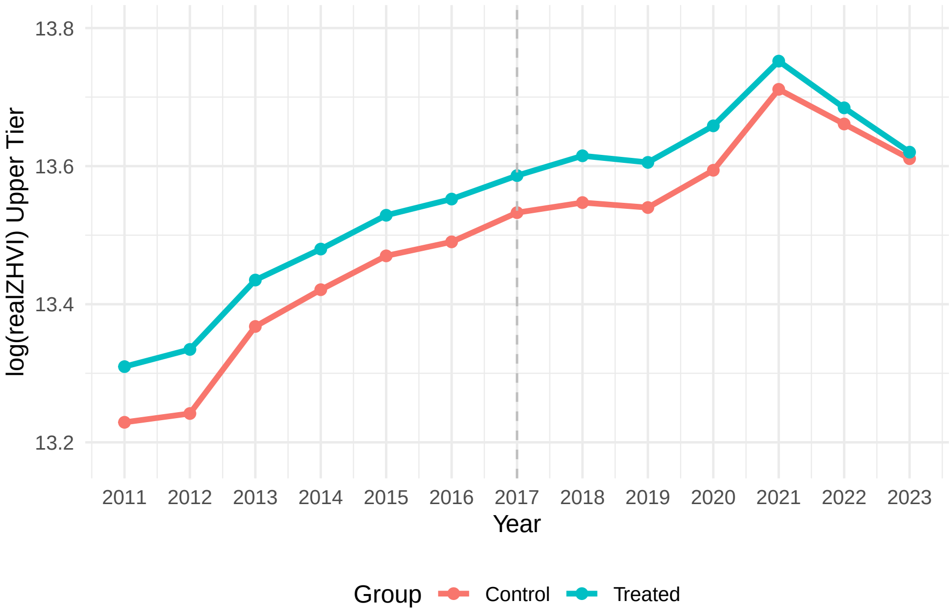

Figure 9: Average Log-transformed Real Zillow Home Value Index (upper tier) by treatment group over time. Dashed line indicates 2017 treatment threshold.

6 Robustness Checks

Individual lag tests

The first approach to assess the robustness of the fixed effects estimates is to individually test the current and lagged effect of heightened wildfire activity on home value indices. Exploring different lags separately helps ensure that the estimated coefficients are not the result of arbitrary modeling choices, but instead reflect the actual dynamics of the effect. Across all tiers, the lags of the burned area share have been tested individually. The significance remains the same for each lag across tiers. Most of the lagged effects show slightly higher magnitudes, which suggests that they partially capture the effect of wildfire exposure from other years that is omitted in single-lag specifications. This supports the interpretation of a persistent, delayed response in housing markets to repeated or recent wildfire activity.

Additionally, different lags of the net migration rate and housing growth rate were tested. For the migration rate, a statistically significant negative effect is found in the same year for the middle and upper tiers. However, the significance disappears once the variable is lagged by one or more years. For housing growth, the effect remains positive throughout, which—as explained earlier—is somewhat unexpected. This suggests that a possible reverse causality mechanism, where people move to counties with lower house prices and construction responds to unmet demand, is more immediate in the same year. In the two-year lag of housing growth becomes statistically significant again. In the lower tier, no significant effect is measured for either the same year or lagged migration rate or housing growth—except for a positive and significant effect of the migration rate in the second-year lag. This could hint at the lower segment of the housing market being more sensitive to migration pressure over longer horizons.

Excluding counties with most severe wildfire activity

The second robustness check focuses on excluding the counties with the highest levels of wildfire exposure. This serves to assess whether the estimated negative impact of wildfires on county-level home values is primarily driven by those counties most severely exposed, or whether a generalisation can be made that also applies to less heavily affected counties. As visible in Figure 3, the counties Plumas, Lake, Trinity, Napa and Butte exhibit the highest levels of wildfire activity throughout the observed period.

As the regression results (Table 5) show, the negative effects measured for the lagged wildfire variables lose their statistical significance across all tiers of the home value distribution once these counties are excluded. The only exception is the contemporaneous effect in the upper tier, where a positive effect of 0.136 % is measured and found to be significant at the 5 % level. When excluding the five counties one by one, the magnitude of the negative effects is reduced in all cases, with Lake and Trinity contributing most strongly to the main result.

This robustness check suggests that it is primarily the counties most severely affected by wildfire activity—measured in terms of the average share of area burned per year—that drive the negative treatment effect identified in the main fixed effects specification. This may imply that the negative supply-side effects described by Boustan et al. (2022), such as the destruction of local amenities, only become sufficiently strong in counties that are hit particularly hard. Alternatively, in the counties remaining in the reduced sample, the potential supply-side effect of a reduced housing stock may partly or fully offset the negative demand-side effects of wildfire exposure. This could even explain the small, positive contemporaneous effect observed in the upper tier.

	Mid Tier	Lower Tier	Upper Tier
PctAreaBurned_t	0.00129	0.00056	0.00126^{*}
	(0.00067)	(0.00089)	(0.00060)
$\text{PctAreaBurned}_{t-1}$	0.00046	-0.00013	0.00022
	(0.00070)	(0.00084)	(0.00065)
$\text{PctAreaBurned}_{t-2}$	-0.00059	-0.00075	-0.00074
	(0.00093)	(0.00109)	(0.00083)
$\text{PctAreaBurned}_{t-3}$	0.00058	-0.00009	-0.00001
	(0.00083)	(0.00079)	(0.00082)
log(AverageRooms)	0.30762^{*}	0.50125^{***}	0.05690
	(0.14622)	(0.14068)	(0.15334)
log(HousingAge)	-0.12620	-0.14369	-0.05729
	(0.06856)	(0.11358)	(0.05932)
UnemploymentRate	-0.01556^{***}	-0.01744^{***}	-0.01162^{***}
	(0.00357)	(0.00358)	(0.00286)
log(RealMedianIncome)	0.34034^{**}	0.38797^{***}	0.28076^{**}
	(0.12705)	(0.11727)	(0.09709)
$\text{NetMigrationRate}_{t-1}$	-0.01387	-0.00461	-0.00835
	(0.01178)	(0.00990)	(0.01000)
$\text{HousingGrowth}_{t-1}$	0.35339	0.39805	0.51763^{*}
	(0.31186)	(0.46602)	(0.24301)
log(MedianAge)	0.33424	0.30845	0.28758
	(0.30868)	(0.28391)	(0.30070)
HighSchoolShare	0.01084^{***}	0.00870^{**}	0.00810^{**}
	(0.00256)	(0.00301)	(0.00245)
AcademicShare	0.00289^{*}	-0.00160	0.00487^{***}
	(0.00113)	(0.00151)	(0.00116)
R-squared	0.26035	0.27380	0.22607
Adj. R-squared	0.16568	0.18001	0.12702
N	51	50	51
T	$12-13$	$12-13$	$12-13$
Number of obs.	662	648	662

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; standard errors in parentheses

Table 5: Regression table on the effects of wildfire activity on different segments of the home value distribution (reduced sample, excluding top 5 counties by average percentage of area burned).

Varying cutoffs for the difference in differences estimation

The cutoff value of 3 % that was chosen for the main DID specification is the highest possible average percentage of area burned in a county per year that still allows for a sound divide between the treatment group and the control group. However, multiple different cutoff values lower than 3 % were implemented to check the robustness of the DID effect (Table 6). Lowering the cutoff to 2.5 % still yields statistically significant negative DID coefficient estimates in all tiers of the home value distribution. For the middle tier, the effect is measured to have a similar, yet slightly stronger magnitude than with the 3 % cutoff, while in the lower tier the measured effect (−4.8 %) is less severe than in the original model. The estimated DID coefficient loses statistical significance when lowering the cutoff to 2 % in the middle and upper tier. In the upper tier, however, the effect remains significant with a 2 % cutoff and is still marginally significant with a 1.5 % cutoff, as its magnitude more than halves to −2.416 %. With a 1 % cutoff, the measurements of the DID coefficient are insignificant across all tiers.

Overall, the DID coefficients decrease in magnitude and lose statistical significance as the cutoff is lowered. This suggests that the effect of wildfire activity on home value indices is better identified and more severe when a higher threshold of average area burned in a county is chosen as a treatment criterion. This is consistent with the measurements of the fixed effects regression, which estimate higher percentages of burned areas to cause a stronger negative effect on home value indices.

Table 6: Robustness Check: Cutoff Values (Threshold Year: 2017)

Cutoff	**Mid Tier**	**Lower Tier**	**Upper Tier**
3 %	-0.03002^{**}	-0.05424^{***}	-0.02246^{*}
2.5 %	-0.03320^{**}	-0.048^{***}	-0.02624^{*}
2 %	-0.01223	-0.02779^{**}	-0.01136
1.5 %	-0.00784	-0.02416^{**}	-0.01049
1 %	-0.0025	-0.01065	-0.00671

Significance levels: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Another possible modification of the quasi-experimental DID estimation is choosing a different threshold year to split the observed time period into two time frames, representing the instances before and after the treatment respectively. While 2017, the year chosen in the main DID specification, marks a year in which wildfire activity rises considerably and stays elevated afterwards, another plausible threshold year is 2020, where an unmatched spike in wildfire activity occurred (Figure 5). With this threshold year, a higher cutoff is possible, since overall wildfire activity is higher on average after 2020 than after 2017. The number of treated counties is 12 when choosing a 4 % cutoff and would drop to only

8 when choosing 4.5 %, which is unfeasible. 4 % is therefore chosen as the highest cutoff of the robustness check.

The caveat of choosing this year is that the two-sample t-test shows a significantly higher unemployment rate in the treatment group, while the median income is significantly lower using the 4 % cutoff. This could cause a downward bias on the estimate of the DID effect. Still, the estimated coefficients (Table 7) across all tiers yield a drastically lower magnitude than in the specification with 2017 chosen as the threshold year, even when comparing the more extreme shock identified with a 4 % cutoff to the one identified with the 3 % cutoff in the main DID specification. Considering that the fixed effects regression estimates a significant decrease in home value indices only two years after elevated wildfire activity in the upper and middle tier, and after one year in the lower tier, the negative impact was expected to be less detectable with the 2020 threshold year, since the post-treatment time period only spans four instances. The lagged effects induced by yearly wildfire shocks from 2020 to 2023 therefore have not had as much time to unfold. For example, the highly significant and negative effects that are measured to occur with a three-year lag are not included in this DID specification, if the wildfire shock took place in 2021 or later. This also points towards the possibility of supply shocks caused by the destruction of housing stock partly offsetting the negative demand-side effects of heightened wildfire activity on home value indices in the short run.

Table 7: Robustness Check: Cutoff Values (Threshold Year: 2020)

Cutoff	**Mid Tier**	**Lower Tier**	**Upper Tier**
4 %	-0.01767^{*}	-0.03265^{*}	-0.00618
3.5 %	-0.01826	-0.02957^{*}	-0.01493^{*}
3 %	-0.01197	-0.02187	-0.01024
2.5 %	-0.01187	-0.02665	-0.01260
2 %	-0.02217	-0.02907	-0.02463^{*}

Significance levels: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

7 Conclusion

The main finding of this thesis is a significant negative effect of heightened wildfire activity on Californian home value indices at the county level, that unfolds with a time lag. The magnitude of the effect is similar in all tiers of the home value distribution over the first three years following an increase in a county's burned area. While in the middle and upper tier of the home value distribution a significant effect is measurable after the second year of the wildfire shock, in the lower tier the significance emerges already after the first year. For both the second and the third year after a wildfire shock, a significant negative effect between -0.2 and -0.3% per additional percentage point of area burned in a county is measured across all tiers of the home value distribution.

The DID estimates support the findings of the main specification, as the measured treatment effects yield statistically significant, negative values across all tiers. The visualizations of the DID effects show a delayed change of trends, setting in after the threshold year of 2017. The clearest emergence of different trends is visible after 2020, especially in the upper and middle tier. Since a change of the threshold year to 2020 did not yield stronger treatment effects, however, this delay is presumed to be caused by the time lag of the negative impact, also estimated in the fixed effects regression, rather than the spike in wildfire activity in 2020.

While some hedonic pricing studies estimate different effects across the pricing distribution suggest a stronger effect on the upper end of the home value distribution, there was no clear difference of the effect detected across the different tiers observed on the more aggregate county level. Although the negative effect of on the upper tier home value indices after three years of a wildfire shock is the strongest across all tiers and lags, the difference in magnitude is not severe. Also, lower tier home value indices show a more immediate significant negative response in the fixed effects model, while this tier also shows consistently stronger estimates of the treatment effect in the DID model than the other tiers.

Notably, an exclusion of the counties most exposed to wildfires over the observed period shows that the measured negative impact is largely driven by the counties experiencing the strongest wildfire shocks, as most coefficients loose significance in the smaller sample. The only significant coefficient estimated in the reduced sample is a small positive effect in upper tier in the county upper tier home value indices of the same year.

Following the theoretical reasoning of Boustan et al. (2020), the negative demand-side effects caused by wildfire activity drive down home values in a county. The relevant channels in this context are firstly the destruction of amenities, prompting a negative effect also captured by numerous hedonic pricing studies, and secondly the destruction of capital stock. The latter is presumed to lead to a lower marginal product of labor and hence lower wages, reducing the net migration and therefore the demand for housing in a county. However, the delayed impact suggests that these effects could be offset temporarily by the destruction of housing stock, which is in generally expected to be of greater relevance in the short term, as housing supply needs time to adjust (Boustan et al. 2020, Hennighausen / James 2024). The erosion of the negative effect in the reduced sample excluding the five counties with the highest wildfire exposure suggests that the negative demand-side effects could be asymmetrically dominant in counties hit very severely. Moreover, spillover effects driving up the demand for housing in the counties surrounding those most impacted by wildfires seem worthy of closer inspection, even if the inclusion of spatial lags into the model was rejected in this thesis by specification tests. Further research is needed to carefully disentangle the proposed channels through which the negative or positive impacts of wildfires on home values operate.

References

Adachi, J.K. and Li, L. (2023). 'The impact of wildfire on property prices: An analysis of the 2015 Sampson Flat Bushfire in South Australia', *Cities*, 136, Article 104255.

American Census Bureau (2024). 'TIGER/Line Shapefiles', June 12, 2025. Available at: https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html

American Census Bureau (2025). 'Using 1-Year or 5-Year American Community Survey Data', June 12, 2025. Available at: https://www.census.gov/programs-surveys/acs/guidance/estimates.html

American Community Survey (2025). 'American Community Survey (ACS)', June 12, 2025. Available at: https://www.census.gov/programs-surveys/acs.html

Anselin, L. (2022). 'Spatial econometrics', in: Rey, S.J. and Franklin, R.S. (eds.), *Handbook of Spatial Analysis in the Social Sciences*, pp. 101–122, Edward Elgar Publishing.

Bivand, R. and Wong, D. (2018). 'Comparing implementations of global and local indicators of spatial association', *TEST*, 27(3), pp. 716–748.

Bogin, A.N., Doerner, W.M. and Larson, W.D. (2019). 'Local house price dynamics: New indices and stylized facts', *Real Estate Economics*, 47(2), pp. 365–398.

Boustan, L.P., Kahn, M.E., Rhode, P.W. and Yanguas, M.L. (2020). 'The effect of natural disasters on economic activity in US counties: A century of data', *Journal of Urban Economics*, 118, Article 103257.

California Department of Forestry & Fire Protection (2025). 'California Fire Perimeters (all)', June 6, 2025. Available at: https://data.ca.gov/dataset/california-fire-perimeters-all

California Department of Insurance (2025). 'Residential Property Insurance Report', June 6, 2025. Available at: https://www.insurance.ca.gov/0400-news/0200-studies-reports/0250-homeowners-study/

Croissant, Y. and Millo, G. (2008). 'Panel data econometrics in R: The plm package', *Journal of Statistical Software*, 27(2), pp. 1–43.

Damianos, D. and Shabman, L. (1976). 'Flood hazard effects on residential property values', *Journal of the Water Resources Planning and Management Division*, 102(1), pp. 151–162.

Driscoll, J.C. and Kraay, A.C. (1998). 'Consistent covariance matrix estimation with spatially dependent panel data', *Review of Economics and Statistics*, 80, pp. 549–560.

Dong, H. (2024). 'Climate change and real estate markets: An empirical study of the impacts of wildfires on home values in California', *Landscape and Urban Planning*, 247, Article 105062.

Donovan, G.H., Champ, P.A. and Butry, D.T. (2007). 'Wildfire risk and housing prices: A case study from Colorado Springs', *Land Economics*, 83(2), pp. 217–233.

Federal Reserve Bank of St. Louis (2025). 'FRED API', June 16, 2025. Available at: https://fred.stlouisfed.org/docs/api/fred/

Greenberg, M., Angelo, H., Losada, E. and Wilmers, C.C. (2024). 'Relational geographies of urban unsustainability: The entanglement of California's housing crisis with WUI growth and climate change', *Proceedings of the National Academy of Sciences*, 121(32), Article e2310080121.

Hansen, W.D. and Naughton, H. (2013). 'The effects of a spruce bark beetle outbreak and wildfires on property values in the wildland-urban interface of south-central Alaska, USA', *Ecological Economics*, 96, pp. 141–154.

Hennighausen, H. and James, A. (2024). 'Catastrophic fires, human displacement, and real estate prices in California', *Journal of Housing Economics*, 66, Article 102023.

Hoechle, D. (2007). 'Robust standard errors for panel regressions with cross-sectional dependence', *The Stata Journal*, 7(3), pp. 281–312.

Internal Revenue Service (2025). 'SOI Tax Stats - Migration Data', July 1, 2025. Available at: https://www.irs.gov/statistics/soi-tax-stats-migration-data

Koo, K.M. and Liang, J. (2022). 'A view to die for? Housing value, wildfire risk, and environmental amenities', *Journal of Environmental Management*, 321, Article 115940.

Loomis, J. (2004). 'Do nearby forest fires cause a reduction in residential property values?', *Journal of Forest Economics*, 10(3), pp. 149–157.

Manely, J. (2013). 'Scale, Aggregation, and the Modifiable Areal Unit Problem', in: Fischer, M.M. and Nijkamp, P. (eds.), *Handbook of Regional Science*, pp. 1157–1171, Springer.

McCoy, S.J. and Walsh, R.P. (2018). 'Wildfire risk, salience & housing demand', *Journal of Environmental Economics and Management*, 91, pp. 203–228.

McGrath, M. (2025). 'Climate 'whiplash' linked to raging LA fires', *BBC News*, 9 January 2025. Available at: https://www.bbc.com/news/articles/c0ewe4p9128o.

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

Mueller, J.M., Loomis, J.B. and González-Cabán, A. (2009). 'Do repeated wildfires change homebuyers' demand for homes in high-risk areas? A hedonic analysis of the short and long-term effects of repeated wildfires on house prices in Southern California', *The Journal of Real Estate Finance and Economics*, 38(2), pp. 155–172.

Nicholls, S. (2019). 'Impacts of environmental disturbances on housing prices: A review of the hedonic pricing literature', *Journal of Environmental Management*, 246, pp. 1–10.

Olsen, C. (2023). 'Zillow Home Value Index Methodology, 2023 Revision: What's Changed?', June 6, 2025. Available at: https://www.zillow.com/research/methodology-neural-zhvi-32128/

Paudel, J. (2021). 'Environmental disasters and property values: Evidence from Nepal's forest fires', *Land Economics*, 98(1), pp. 115–131.

Pebesma, E. (2018). 'Simple Features for R: Standardized Support for Spatial Vector Data', *The R Journal*, 10(1), pp. 439–446.

Pricewaterhouse Coopers (2001). *Economic Study of the Los Alamos*, June 6, 2025. Available at: https://www.fema.gov/news-release/2001/03/21/economic-study-los-alamos

Rossi, D. and Byrne, D.A. (2016). 'Spatially weighted hedonic property models for homes vulnerable to wildfire', in: *Proceedings of the 5th International Fire Behavior and Fuels Conference*, Portland, Oregon, USA, April 11–15, 2016.

Stetler, K.M., Venn, T.J. and Calkin, D.E. (2010). 'The effects of wildfire and environmental amenities on property values in northwest Montana, USA', *Ecological Economics*, 69, pp. 2233–2243.

Swain, D. L., Prein, A. F., Abatzoglou, J. T., Albano, C. M., Brunner, M., Diffenbaugh, N. S., Singh, D., Skinner, C. B. and Touma, D. (2025). 'Hydroclimate volatility on a warming Earth', *Nature Reviews Earth Environment*, 6(1), pp. 35–50.

Zillow (2025). 'Housing Data', June 10, 2025. Available at: https://www.zillow.com/research/data/

8 Appendix

8.1 Tests for spatial specifications

Table 8: Summary of the p-values of Lagrange multiplier tests for different spatial specifications, for OLS regressions of each time period.

Year	RSlag	RSerr	adjRSlag	adjRSerr	SARMA
2011	0.0228	0.1981	0.0601	0.9264	0.0746
2012	0.2113	0.3074	0.4032	0.6729	0.4189
2013	0.1082	0.8039	0.0474	0.2347	0.1359
2014	0.0868	0.6022	0.0247	0.1224	0.0700
2015	0.4035	0.6396	0.2021	0.2838	0.3972
2016	0.4235	0.3791	0.7105	0.6027	0.6340
2017	0.3781	0.6583	0.4443	0.9518	0.6769
2018	0.0668	0.6476	0.0572	0.4948	0.1477
2019	0.0795	0.5515	0.0940	0.7709	0.2060
2020	0.0184	0.0660	0.1155	0.5839	0.0534
2021	0.1361	0.5137	0.1653	0.7178	0.3086
2022	0.4093	0.5044	0.5816	0.7936	0.6875
2023	0.7518	0.9710	0.6865	0.8000	0.9212

As explained in the methodology section, none of the tested spatial specifications consistently perform better than the reduced model. None of the set-ups to include spatially lagged variables or error terms adequately capture spatial dependence throughout the time period, as the table of p-values shows (Table 8). However, in some years, a significant difference of the reduced model and the RSlag or the adjRSlag model respectively are found at a 5 % level. To consider possible spatial dependence in the statistical inference conducted in this thesis, Driscoll-Kraay standard errors are introduced.

8.2 Effect of wildfire activity on average insurance premiums

The effect of wildfire activity on real estate markets is investigated in a separate fixed effects regression (Table 9), employing the same explanatory variables. The results show a statistically significant and positive effect of lagged burned area on insurance premiums after the first year and persisting in the third year after the shock. This aligns with expectations: higher wildfire activity likely increases perceived and actual risk, raising premiums. The small effect sizes and the lack of significance in the contemporaneous and third lag may reflect institutional constraints. For instance, government regulations may prevent insurers from rapidly adjusting premiums, or insurers may exit high-risk markets entirely, as noted by Powell at al. (2023).

	log(RealAveragePremium)
PctAreaBurned_t	0.00027
	(0.00044)
$\text{PctAreaBurned}_{t-1}$	0.00059^{*}
	(0.00026)
$\text{PctAreaBurned}_{t-2}$	0.00104^{**}
	(0.00036)
$\text{PctAreaBurned}_{t-3}$	0.00015
	(0.00051)
log(AverageRooms)	0.12552
	(0.08223)
log(HousingAge)	0.30264^{***}
	(0.05311)
UnemploymentRate	-0.00024
	(0.00132)
log(RealMedianIncome)	-0.03889
	(0.02120)
$\text{NetMigrationRate}_{t-1}$	0.00507
	(0.00401)
$\text{HousingGrowth}_{t-1}$	-0.17186^{*}
	(0.08399)
log(MedianAge)	-0.41553^{***}
	(0.05255)
HighSchoolShare	0.00184^{**}
	(0.00067)
AcademicShare	0.00243
	(0.00151)
R-squared	0.13066
Adj. R-squared	-0.10905
N	58
T	6
Number of obs.	348

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; standard errors in parentheses

Table 9: Regression table on the effects of wildfire activity on the average insurance premium.

The analysis is limited by the short time frame of insurance data (2018–2023) and the relatively small panel (348 observations). The model has low explanatory power ($R^2 = 0.13$, adjusted $R^2 = -0.11$). This mainly results from the small sample size and the inclusion of year fixed effects, which absorb much of the temporal variation. Excluding these fixed effects results in a substantial increase in both R^2 values, suggesting that much of the variation in premiums occurs over time rather than across counties. Again, prioritizing the identification of a significant effect rather than increasing predictory power, an omission of year fixed effects is not feasible.

Because of these problems, these findings should therefore be interpreted with a lot of caution. Extending the panel to cover a longer time period would likely improve the robustness of the estimates and allow for better identification of the dynamic effects of wildfire exposure on insurance pricing. This is not possible, as insurance data on a county level is provided only since 2018, as explained in the methodology section.