A Additional Results and Robustness Checks

A.1 Effect of Homes on Fire Costs

A.1.1 Robustness Checks

Appendix Table 1 shows the results from Table 1 in the main text, including coefficients on the control variables as well as an additional "no controls" specification. It also shows an additional specification that includes controls for the distance from the ignition point to the nearest primary road.[30]

Appendix Table 2 shows a robustness check proposed by Oster (2019), building on Altonji, Elder, and Taber (2005) and related work. This sensitivity test bounds the potential bias from unobservable confounders under an assumption about δ, the relative degree of selection on observables and unobservables, and an assumption about R_{max}, the R^2 of a hypothetical regression containing all the observables and unobservables. Oster (2019) shows that for $\delta = 1$ (equal selection on onbservables and unobservables), the bias-adjusted treatment effect β^* is approximately $\tilde{\beta} - [\dot{\beta} - \tilde{\beta}]\frac{R_{max}-\tilde{R}}{\tilde{R}-\dot{R}}$. Here, $\tilde{\beta}$ and $\tilde{R}$ are the coefficient and R^2 from a regression with the full set of controls, and $\dot{\beta}$ and $\dot{R}$ come from a restricted specification. This approximate formula provides intuition: results are more robust when including controls produces smaller changes in the coefficient, and larger increases in the R^2.

We implement the exact version of the calculation provided in Oster (2019) and the software package *psacalc*. Because Oster's test is limited to a scalar treatment, we implement the regression test for a linear version of Equation 3, where $g(Homes)$ is the distance from the ignition point to the nearest home (this is a mild restriction given the near-linearity apparent in Figure 3). The restricted specification includes only national forest fixed effects. The controlled specification is Column (3) from Table 1, the richest set of controls that we discuss. It includes the weather, topography, and vegetation variables described in Table 1 and Appendix Table 1. It also includes year-month by state (i.e., month of sample by state) dummies that proxy for unobservable changes in fire risk due to factors such as fuel dryness. We follow Oster (2019) and assume that $R_{max} = 1.3\tilde{R}$. The final column of Appendix Table 2 reports Oster's recommended quantity, an "identified set" for the effect of distance to homes on fire costs. The lower bound is the bias-adjusted treatment effect assuming δ=1, and the upper bound is $\tilde{\beta}$. In Oster's framework, results are considered robust when this set excludes zero. This condition holds in our case. Furthermore, the lower bound on the treatment effect of -0.042 is similar to the fully-controlled regression coefficient of -0.050.

Appendix Table 3 shows additional robustness checks for the effects of the number of

30. Road data come from the US Census TIGER/Line shapefile for primary roads for 2016. Primary roads roughly correspond to interstate highways.

nearby homes on fire costs. Columns (1) through (5) show the same checks that we showed for the effect of the nearest home in Table 1. Our results are robust to these various tests. Column (6) shows an additional specification that measures the stock of nearby homes by total transaction value, instead of number of homes. Results are again similar.

Appendix Figure 1 shows results using different radii around the ignition point to count threatened homes. The omitted category in each regression is fires with zero homes within the radius. The other bins in each regression are defined by deciles of number of homes, conditional on any homes within the radius. For all three radii, there is a clear pattern of quick increases across the first two bins, and then roughly constant costs at higher numbers of homes. Note that direct comparisons of these coefficients across bins are difficult, since the comparison group of fires with zero threatened homes is systematically different across columns (e.g., for 40 km, all fires with zero homes are very remote by construction). Several other effects also presumably occur simultaneously as we widen the radius: since further-away homes have less effect on costs, these measures attenuate somewhat; however, because calculating density over a wider area may reduce noise in our assessment of the number of threatened homes, there may be another factor making these measurements more precise. Finally, note that the actual bin endpoints vary across models. Importantly, however, the obvious non-linear pattern of costs by number of homes exists for any radius.

Appendix Figure 2 plots covariate overlap for the covariates included in the regressions.

Appendix Table 1: The Effect of Proximity to Homes: Full Results

	(1)	(2)	(3)
10–20 km	-0.5232	-0.3436	-0.4099
	(0.1709)	(0.1491)	(0.1586)
20–30 km	-1.1075	-0.9018	-0.9957
	(0.3261)	(0.2676)	(0.2861)
30–40 km	-2.4784	-1.6605	-1.7601
	(0.3796)	(0.4528)	(0.5275)
40+ km	-2.7290	-2.0257	-2.1063
	(0.3631)	(0.3774)	(0.4511)
WindSpeed		0.0642	0.0691
		(0.0347)	(0.0347)
WindSpeed^2		-0.0017	-0.0019
		(0.0013)	(0.0013)
TerrainSlope		0.0413	0.0420
		(0.0181)	(0.0185)
TerrainSlope^2		-0.0007	-0.0007
		(0.0004)	(0.0004)
VaporPressureDeficit		0.0681	0.0661
		(0.0371)	(0.0351)
$\text{VaporPressureDeficit}^2$		-0.0015	-0.0014
		(0.0007)	(0.0007)
Precipitation		-0.0513	-0.0446
		(0.0440)	(0.0432)
Precipitation^2		0.0010	0.0010
		(0.0010)	(0.0010)
South/SW Aspect		0.2361	0.2322
		(0.1356)	(0.1362)
Shrub Fuel Model		-0.1266	-0.1482
		(0.1926)	(0.1921)
Timber Fuel Model		-0.0829	-0.0900
		(0.1545)	(0.1527)
Slash Fuel Model		0.5048	0.4466
		(0.3638)	(0.3730)
Urban/Barren Fuel Model		-0.1804	-0.1837
		(0.2460)	(0.2476)
Distance to Primary Road			0.0109
			(0.0064)
$(\text{Distance to Primary Road})^2$			-0.0000
			(0.0000)
Constant	13.5168	10.8345	10.1655
	(0.1873)	(1.6243)	(1.6171)
National Forest FE		X	X
Year by State FE		X	X
Month-of-Year by State FE		X	X
Fires	2,089	2,089	2,089
R^2	0.09	0.43	0.43

Notes: Column (2) reproduces Column (2) of Table 1, showing coefficients for the controls. Column (1) shows a no-controls specification for comparison. Terrain slope is the linear slope of the ground surface. Wind speed is average speed on the day of ignition at the reference weather station listed in FAMWEB (in miles per hour). Vapor pressure deficit is for the ignition location and day, from PRISM, and measured in hectopascals (millibars). Precipitation is the amount of precipitation on the ignition day in mm, from PRISM. Fuel model fixed effects include four categories corresponding to LANDFIRE fuel models for brush, grass, timber, and barren/urban/other. The omitted fuel model category is grass. Forest unit fixed effects include 86 national forests in the western US Standard errors are clustered at the national forest level.

Appendix Table 2: Oster's (2019) Coefficient Stability Test

	Restricted Specification	Controlled Specification	Identified Set
Coefficient	-0.054	-0.050	(-0.042,-0.050)
Standard Error	0.005	0.007	
R^2	0.27	0.54	
Included Controls	National Forest FEs	National Forest FEs, Weather, Topography, Vegetation, Month-of-sample by state dummies	

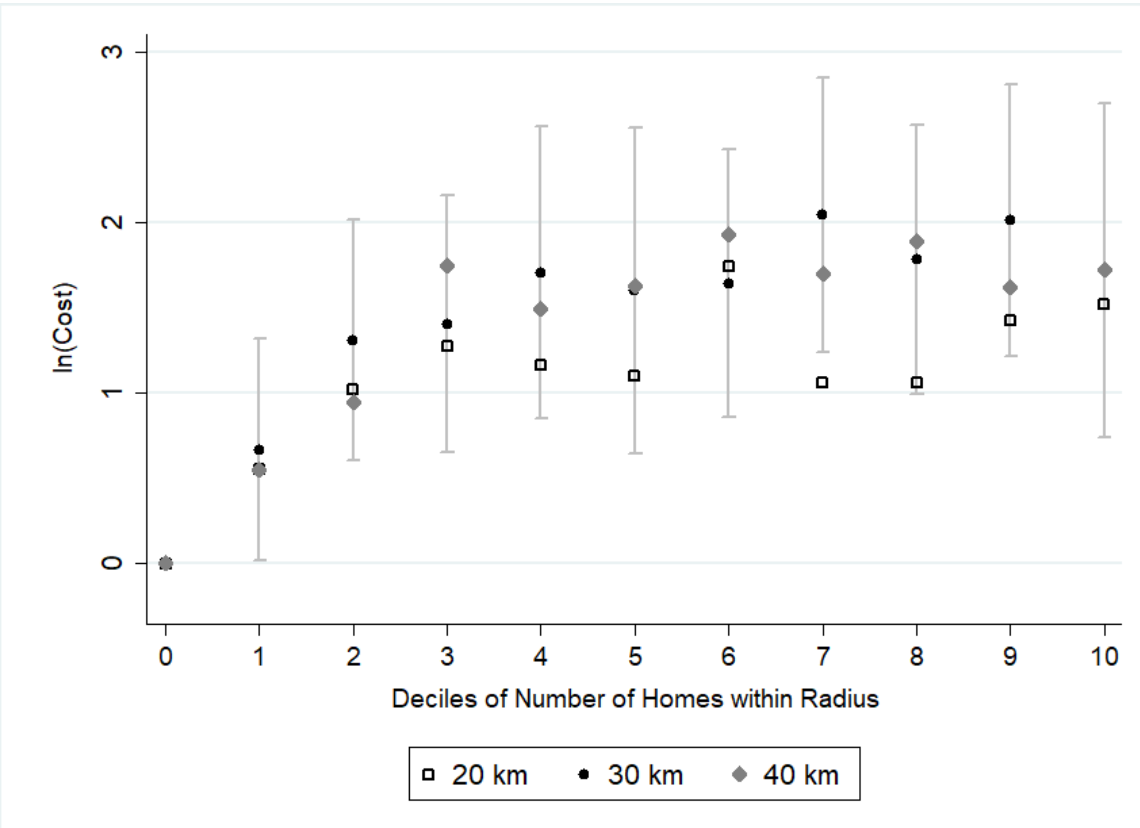
Notes: This table implements a procedure proposed by Oster (2019) to bound selection bias due to unobservable confounders. See the text of Appendix A.1.1 for details.

Appendix Table 3: The Effect of Number or Value of Homes, Robustness Checks

	Number					Value
	(1)	(2)	(3)	(4)	(5)	(6)
Quintile Bins						
1	0.97	0.94	0.91	1.00	1.15	0.89
	(0.31)	(0.31)	(0.36)	(0.34)	(0.69)	(0.32)
2	1.52	1.46	1.38	1.46	1.41	1.43
	(0.38)	(0.37)	(0.40)	(0.39)	(0.54)	(0.42)
3	1.61	1.57	1.37	1.45	1.91	1.64
	(0.45)	(0.43)	(0.48)	(0.45)	(0.66)	(0.40)
4	1.85	1.78	1.75	1.71	2.31	1.86
	(0.39)	(0.37)	(0.44)	(0.43)	(0.65)	(0.37)
5	1.87	1.81	1.54	1.75	1.98	1.83
	(0.43)	(0.41)	(0.47)	(0.49)	(0.70)	(0.41)
Controls for Weather, Topography, and Vegetation		X	X	X	X	X
National Forest FE	X	X	X	X	X	X
Month-of-Year by State FE	X	X		X	X	X
Year by State FE	X	X		X	X	X
Month-of-Sample by State FE			X			
Lightning fires only				X		
Timber Fuels only					X	
N	2,089	2,089	2,089	1,470	772	2,089
R^2	0.42	0.43	0.54	0.45	0.57	0.43

Notes: Columns (1) through (5) reproduce estimates from Figure 4 in the main text, using bins of the number of homes within 30 kilometers as the variables of interest. The bins are equal observation bins for fires with at least one nearby home (see Figure 4 for bin ranges). The omitted category is fires with zero nearby homes. Column (6) shows an alternative specification that measures the stock of homes within 30 km by total transaction value. Again, bins are equal observation bins for fires with at least one nearby home, and the excluded category is fires with zero nearby homes. Homes with unusable transaction values, as defined in Section B.2, are assigned the average transaction value of other homes withing 30 km of the ignition point. See Table 1 for details on controls for weather, topography, and vegetation. Standard errors are clustered by national forest.

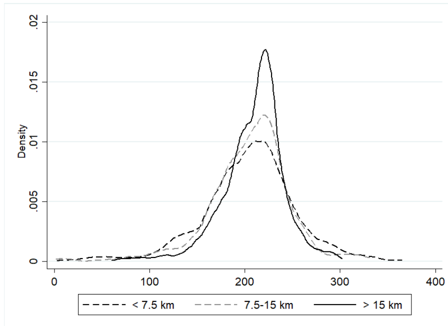
Appendix Figure 1: Costs by Number of Homes: Alternative Radii



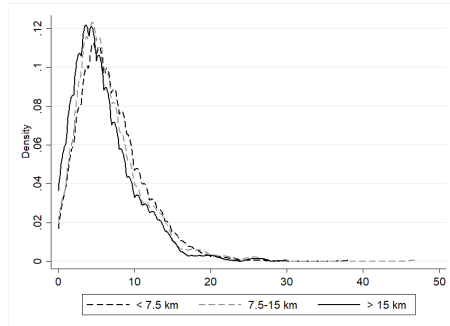
Notes: This figure reproduces Figure 4 from the main text using alternative radii. Each set of markers shows coefficients from a single regression using a different radius around the ignition point of the fire. The bins correspond to deciles of the distribution of number of homes within the radius, conditional on any homes within the radius. The omitted category in each regression is fires with zero homes within the radius. For all three radii, there is a clear pattern of quick increases across the first three to four bins, and then roughly constant costs at higher numbers of homes.

Appendix Figure 2: Covariate Overlap by Distance from Ignition Point to Nearest Home

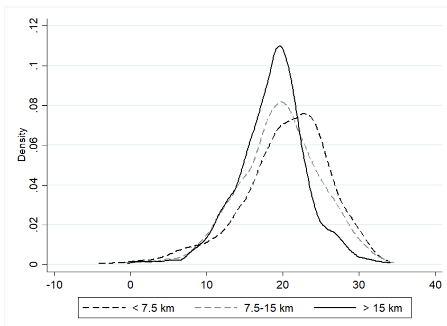
(a) Day of Year (Ignition)



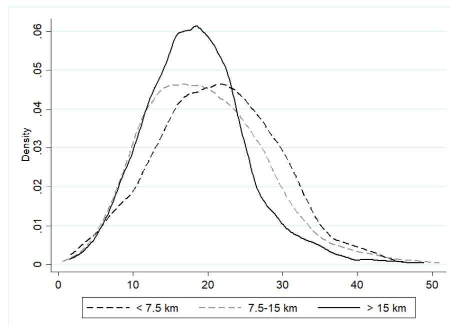
(b) Wind Speed (mph)



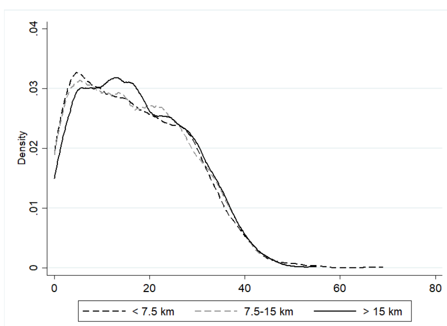
(c) Temperature (F)



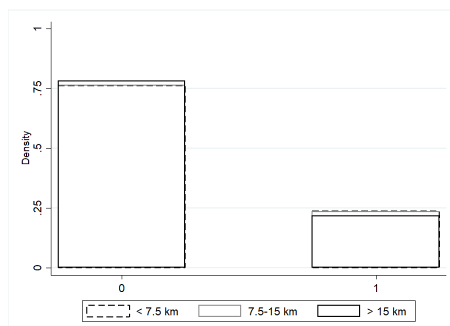
(d) Vapor Pressure Deficit



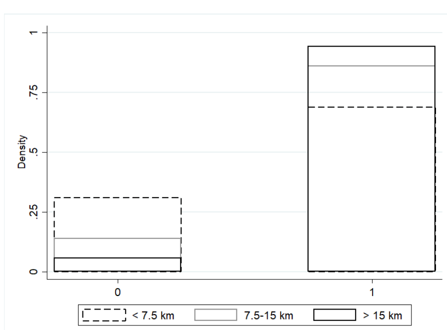
(e) Terrain Slope



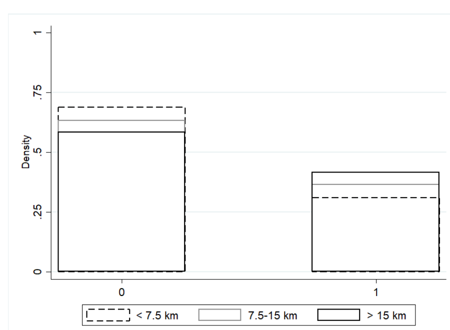
(f) South/southwest-facing



(g) Lightning-caused



(h) "Timber" fuel model



Notes: Figure shows covariate distributions for the US Forest Service fires analyzed in Table 1 and Figures 3 and 4. Panels (b), (c), and (d) report weather on the day of ignition. Wind speed is average wind speed from the reference weather station reported in FAMWEB. Temperature and vapor pressure deficit are mean daily values from PRISM. Terrain slope is the slope percentage, where 100 corresponds to a slope of 1 (i.e, a 45-degree line). "Timber" fuel models follow the Anderson Fire Behavior Fuel Models.

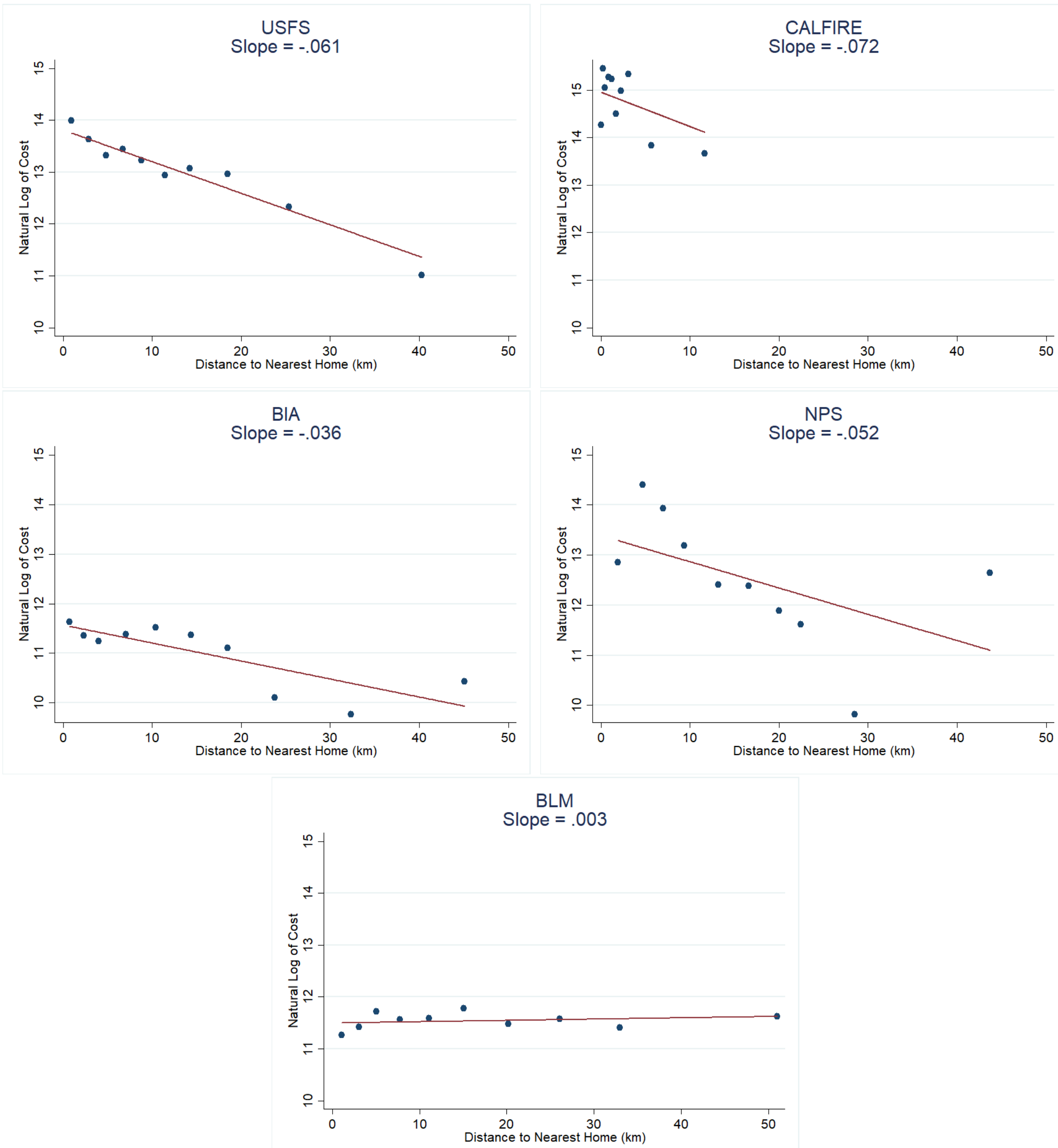
A.1.2 Non-USFS Agencies

The analysis of the effect of home construction on firefighting costs in Section 4 focuses on fires managed by the US Forest Service. Forest Service fires represent the largest group of expenditures and longest time series in our dataset. The national forests also provide a useful source of identifying variation, in that each national forest represents a mostly-contiguous area of public land with broadly similar landscapes and vegetation. This contiguity allows us to take advantage of variation in ignition locations within each of these 86 units using a fixed effects strategy. In comparison, Bureau of Land Management lands are less likely to consist of large contiguous units of land (instead, patches of BLM land in each state are managed by a system of district offices). Similarly, Cal Fire incidents take place on diffuse private and state lands throughout California.

For completeness, this section shows the relationship between homes and ignition costs for each of the agencies from which we were able to obtain data. Given that the empirical design used in the main text is not available for these other agencies, we focus on raw correlations. Appendix Figure 3 plots log firefighting costs against the distance from the ignition point to the nearest home. Across agencies, costs decline for fires located further from homes. Given that the data represent independent administrative databases compiled separately by each agency, the broad similarities across agencies are notable. For the US Forest Service, Cal Fire, the Bureau of Indian Affairs, and the National Park Service, there is a clear downward relationship with a linear slope between -0.036 and -0.073. Bureau of Land Management incidents show a different relationship, with a slope near zero and a lower intercept. One possible explanation for this difference is that it may reflect the characteristics of fires managed by BLM. Compared to USFS fires, the fires managed by BLM are more likely to occur in easier-to-manage grass areas, and less likely to occur in timber fuels. Notwithstanding this pattern for BLM, the broad agreement across the other four agencies is reassuring. This is particularly true given the relatively small size of BLM expenditures relative to USFS and Cal Fire, both overall and in per-incident terms (see Appendix Table 7).

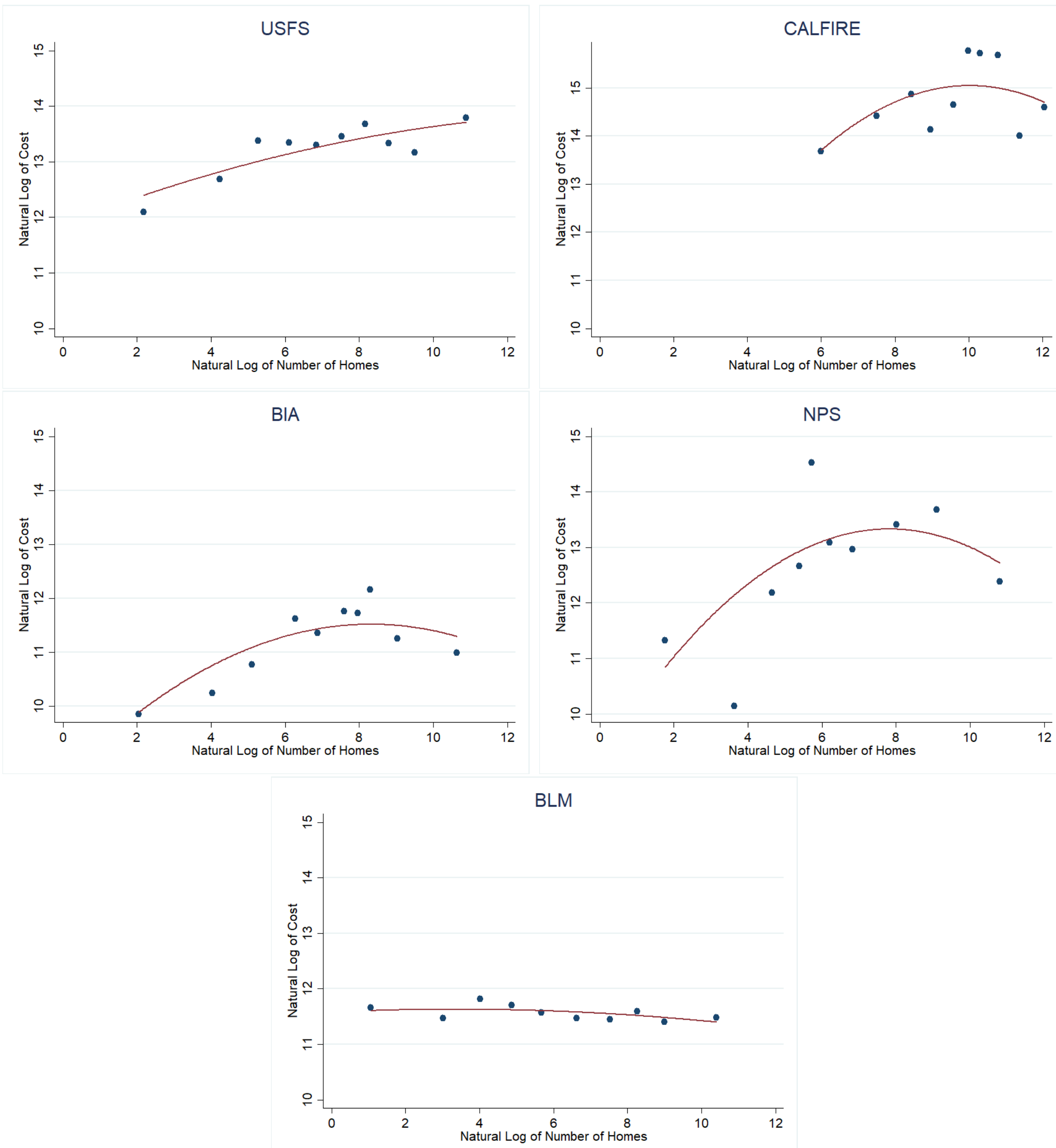
Appendix Figure 4 plots log firefighting costs against the total number of nearby homes. Across agencies, these ln-ln plots imply small or near-zero increases in fire-fighting costs as the number of nearby homes grows large.

Appendix Figure 3: Cost vs. Distance to Nearest Home, by Agency



Notes: Figure shows binned scatterplots for each agency from which we obtained incident expenditure data. The dots show average log incident costs for each decile of distance to nearest home. The red lines show a linear fit. Cal Fire is the California Department of Forestry and Fire Protection; BIA is the Bureau of Indian Affairs; BLM is the Bureau of Land Management; and NPS is the National Park Service.

Appendix Figure 4: Cost vs. Number of Nearby Homes, by Agency



Notes: Figure shows binned scatterplots for each agency from which we obtained incident expenditure data. The dots show average log incident costs for each decile of log number of nearby homes (fires with zero nearby homes are not plotted). The red lines show a quadratic fit. Cal Fire is the California Department of Forestry and Fire Protection; BIA is the Bureau of Indian Affairs; BLM is the Bureau of Land Management; and NPS is the National Park Service.

A.1.3 Effect of Homes on the Number of Fires

To evaluate whether adding homes increases the number of fires (in addition to increasing expenses on each fire), we use panel variation in home construction near national forests in our dataset. We construct a year-by-national forest panel including 76 national forests and 20 years. Because new homes are most likely to affect the number of ignitions in places with relatively low levels of existing development, we exclude national forests with more than 100,000 homes within 30 kilometers of the national forest boundary in 1995 (this excludes 20% of national forest areas with the highest 1995 populations).

We implement a range of panel regression specifications. The outcome variable is the number of fires larger than 300 acres in each forest-year. Our preferred statistical approach is a Poisson regression, since the number of large fires is a count variable.[31] The key identification challenge in this setting is to separate the effect of new home construction from other time-varying determinants of fire probability. Because homes are durable, the number of homes near each national forest increases monotonically across the sample. We adopt a variety of time trends and year fixed effects specifications to control as flexibly as possible for potential secular trends in the number of fires in each national forest caused by factors like climate change or annual drought cycles. Our results in this section should be interpreted with caution, since they rest on the assumption that, conditional on these controls, the trend in new home construction near each national forest is uncorrelated with other trends in fire occurrence.

Appendix Table 4 shows the results. All of these regressions include national forest fixed effects to account for time-invariant determinants of fire risk, such as local topography. Across specifications, new home development has a small positive effect on the number of large fires each year. In Column (1), the estimated coefficient in the Poisson regression is 0.042. This implies that adding 1,000 new homes increases the annual number of fires in this national forest by about 4.3%. The mean number of large fires in each national forest-year is 1.48, so this implies that an additional 1,000 homes lead to 0.06 additional large fires per year. Columns (2)–(5) include alternative polynomial time trends and find similar results. Column (6) instead includes year fixed effects, which allows for arbitrary annual trends at the West-wide level. Column (7) shows the same fixed effects specification in an OLS regression.

31. We use a cluster-robust variance estimator to eliminate the typical limitation of classical Poisson regression, which is that that the mean and variance of the estimates must be equal.

Appendix Table 4: The Effect of Homes on the Number of Fires

	(1) Poisson	(2) Poisson	(3) Poisson	(4) Poisson	(5) Poisson	(6) Poisson	(7) OLS
Thousands of Homes	0.042 (0.008)	0.050 (0.011)	0.040 (0.013)	0.051 (0.011)	0.043 (0.012)	0.040 (0.013)	0.033 (0.018)
National Forest FE	X	X	X	X	X	X	X
Linear Time Trend		X					
Quadratic Time Trend			X				
Regional Linear Trends				X			
Regional Quadratic Trends					X		
Year Fixed Effects						X	X
N	1,180	1,180	1,180	1,180	1,180	1,180	1,180

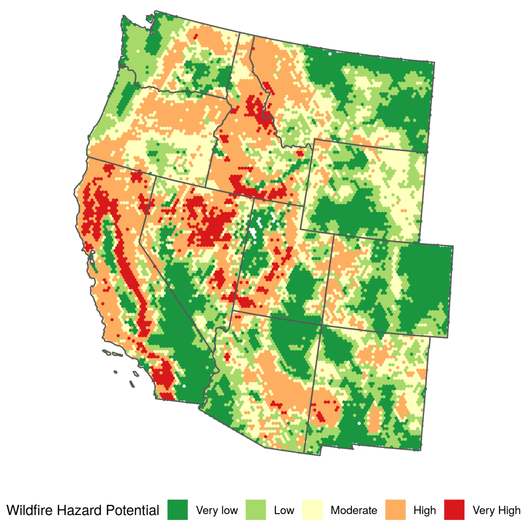
Notes: Table reports the results of seven separate regressions. In each regression the dependent variable is the number of fires larger than 300 acres in each national forest-year. Columns (1)-(6) show results for several Poisson regression specifications, and Column (7) shows an OLS specification for comparison. The variable of interest is the number homes within 30 kilometers of the national forest boundary, in thousands. The table reports regression coefficients and standard errors, which are calculated using a cluster robust variance estimator at the national forest level. For the Poisson specifications, the coefficients can be converted to expected percentage changes in the number of large fires using calculation $e^{\beta} - 1$. See text for details. The mean number of fires in each national forest-year is 1.5. "Regional Linear Trends" and "Regional Quadratic Trends" indicate that the regression includes separate polynomial time trends for each of the five forest service regions included in the sample area.

A.2 Expected Protection Costs

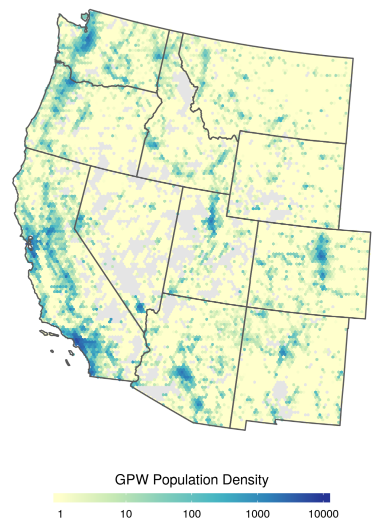
A.2.1 Variables Used to Define Actuarial Groups

Appendix Figure 5: Variables Used to Define Actuarial Groups

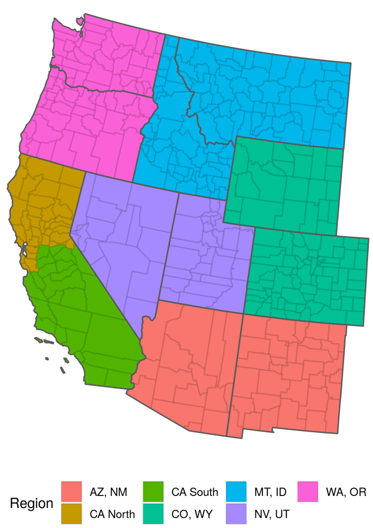
(a) Wildfire Hazard Potential



(b) Population Density



(c) Region



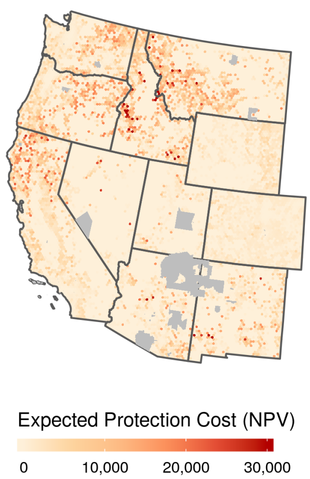
Notes: Wildfire hazard potential: Dillon (2015). Population density: CIESIN (2017).

A.2.2 Maps of Suppression-Only and California Measures

Appendix Figure 6 reproduces the map in Figure 6 using the alternative measures of expected protection cost described in Section 5.1.3 of the main text. Panel A uses the Suppression Only cost measure and Panel B uses the California-specific cost measure. The California measure is displayed as zero for all areas outside California.

Appendix Figure 6: Expected Protection Cost, Alternative Measures

(a) "Suppression Only"



(b) California-specific

Notes: This figure reproduces Figure 6 showing alternative measures of expected protection cost. See Section 5 for a detailed description of the construction of these measures. Units for the color scale are 2017 dollars per home. The California-specific measure in Panel (b) is displayed as zero for areas outside California.

A.2.3 Alternative Measures Based on Interview Evidence

Table 5 compares implicit subsidy estimates using different methods to measure the share of expenditures devoted to protecting homes. Columns 1A, 2A, and 3A show the main estimates from Table 2. Spending on home protection for each incident is the difference between observed costs and predicted costs for that fire in the absence of nearby homes, using the regression model in Section 4. Columns 1B, 2B, and 3B compute analogous subsidy estimates under the alternative assumption that 72.5% of all fire costs are attributable to protecting homes, based on USDA (2006). Comparing 1A to 1B, 2A to 2B, and 3A to 3B shows relatively small differences.

Appendix Table 5: Expected Parcel Protection Costs, Alternative Estimates

	Federal Suppression Only ($)		Suppression Plus ($)		California Only ($)	
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)
Mean	1,077	932	2,408	2,129	2,712	2,315
p50	500	400	1,200	1,100	1,300	1,100
p90	2,100	1,800	5,200	4,500	6,600	5,500
p95	3,800	3,400	8,400	7,300	9,000	7,800
p99	12,700	11,100	22,700	20,900	18,200	15,700
N	8,633,554	8,633,554	8,633,554	8,633,554	3,483,715	3,483,715

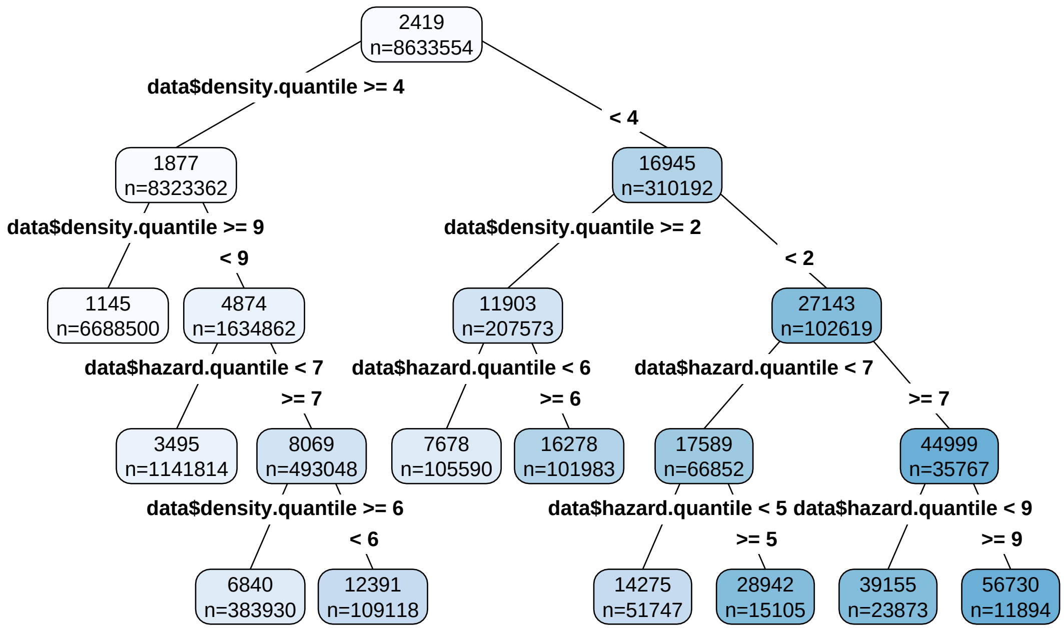
Notes: Columns 1A, 2A, and 3A are identical to Table 2. Columns 1B, 2B, and 3B assume that 72.5% of all fire costs are attributable to homes. The method used to divide protection expenditures across individual homes is the same as in the main analysis.

A.2.4 Machine Learning to Define Actuarial Groups

The main analysis assigns homes to actuarial groups and then averages historical costs for homes in each group to yield expected protection costs. Instead of having the researcher define these actuarial groups, it is possible to use a machine learning technique to define groups. To evaluate the robustness of the actuarial groups used in the main text, we implemented such an approach using a regression tree. Using right-hand-side variables supplied by the researcher, the regression tree algorithm groups homes in order to minimize the prediction error for historical firefighting costs in each group. The number of groups is governed by a complexity parameter that specifies the minimum required improvement in prediction accuracy to justify additional splits. Appendix Figure 7 illustrates the approach. For this figure, we use a high value for the complexity parameter so that there are relatively few splits in the tree. The right-hand-side variables are 10 bins of wildfire hazard potential (WHP) and 10 bins of development density as predictors.

To compare results using this approach to those displayed in Table 2 in the main text, Appendix Table 6 shows the distribution of expected protection costs with a more complex tree using 10 bins of WHP, 10 bins of development density, and the 7 firefighting dispatch regions. This tree generates 79 actuarial groups. The overall distribution of expected protection costs with the regression tree (column 2) is similar to the distribution of expected protection costs in the main analysis (column 1), and

Appendix Figure 7: Illustrative Regression Tree for Defining Actuarial Groups



Notes: This figure illustrates the regression tree approach to defining actuarial groups using a restricted set of predictors and a limited complexity parameter. The top number in each node is the predicted protection cost. The number of homes in each group is given as "n".

the correlation of individual protection costs between the two approaches is 0.8.

Appendix Table 6: Expected Protection Costs using Regression Trees

	(1)	(2)
Mean	2,408	2,416
p50	1,200	1,300
p90	5,200	4,700
p95	8,400	8,000
p99	22,700	22,300
N	8,633,554	8,633,554

Notes: This table shows expected protection costs for the Suppression Plus cost metric. Column 1 is identical to Table 2. Column 2 shows the distribution of costs when actuarial groups are selected using a regression tree algorithm. Percentile cutoffs are rounded to the nearest $100.

A.3 Theory Appendix

Proof that Per-Capita Disaster Costs Decrease with Population

Claim: Per-capita disaster-related costs $\frac{f(n_r)}{n_r} + H(f(n_r))$ decrease with n_r. *Proof*: Take the derivative with respect to n_r and re-arrange.

$$\frac{f'(n_r)}{n_r}\big(1 + n_r H'(f)\big) - \frac{f(n_r)}{n_r^2} \tag{4}$$

Recall f is chosen to minimize $f + n_r H(f)$, so that the derivative $1 + n_r H'(f)$ equals zero. Expression 4 reduces to $-\frac{f(n_r)}{n_r^2}$, which is negative.

Intensive margin changes in risky place population

This section considers the marginal protection cost and the marginal welfare impact of changes in the risky place population. Differentiating Expression 1 with respect to n_r yields the change in net benefits,

$$\theta_{n_r} - s(n_r) - \phi f'(n_r) - \phi\left[H(f(n_r)) + \frac{\partial H}{\partial f(n_r)} f'(n_r) n_r\right] \tag{5}$$

The first term is WTP of the marginal risky place resident. The second is the marginal cost of supplying housing. The third is the expected marginal increase in defensive expenditures. The final term in brackets is the change in expected property damages, which includes expected damages for one more home and decreased expected losses for all inframarginal homes due to increased defensive expenditures during a disaster.

The assumptions in this model allow us to apply the envelope theorem to further simplify Expression 5 to $\theta_{n_r} - s(n_r) - \phi H(f(n_r))$.[32] Compare this expression for social marginal benefit to the private marginal benefit for risky place residents, $\theta_{n_r} - s(n_r) - \phi H(f(n_r)) - \phi \frac{\partial H}{\partial f(n_r)} f'(n_r) n_r$. The latter expression includes an additional term equalling the benefit to inframarginal residents. Thus, private marginal benefit in the risky place exceeds social marginal benefit (recall that $\frac{\partial H}{\partial f} < 0$).

Welfare analysis on the intensive margin depends on assumptions about how development is coordinated. If we assume the marginal resident internalizes all costs and benefits of their location decision except central government expenditures, then failure to price marginal defensive expenditures leads to excess development in the risky place. Such an assumption may be justified if a local government manages risky place development to maximize local benefits, or if risky place residents arrange private side payments. If we instead assume that the marginal resident receives no compensation

32. Rewrite $\theta_{n_r} - s(n_r) - \phi H(f(n_r)) - \phi f'(n_r)[1 + \frac{\partial H}{\partial f} n_r]$. Optimality of f means that $1 + \frac{\partial H}{\partial f} n_r = 0$.

for benefits to inframarginal households, then the failure to price marginal defensive expenditures is offset by this second externality. If dispatch of defensive expenditures during disasters is exactly optimal and we only consider small changes in population, these externalities offset exactly and providing defensive expenditures for free yields the optimal result on the intensive margin.

Let us step back from this ambiguous result and consider the empirical analysis. We find that $f'(n_r)$ is near zero in already-developed areas. This means that any intensive margin distortion due to subsidized marginal protection costs would be small. It also means that spillover benefits to inframarginal residents must be small because there is little actual change in firefighting dispatch. Thus, regardless of what one assumes about how development proceeds in already-developed places, our results imply that any intensive margin distortions are small. What matters for welfare is instead new development in undeveloped and sparsely-developed high-risk places, where the large average protection costs that we measure imply that *total* benefits may not exceed *total* social cost.

Extending the Theoretical Model to Private Self-protection

Let g represent the amount of private risk-reducing investment by each identical homeowner in the risky place. Private damages in the event of a disaster are now $H(f,g)$, with $\frac{\partial H}{\partial g} < 0$ and $\frac{\partial^2 H}{\partial g^2} > 0$. Assume that the central government takes g and n_r as given when choosing f during a disaster (as happens for wildfire and other natural hazards). The optimal emergency defensive expenditure f during a disaster is now given by,

$$f^*(n_r, g) = \arg\min_f f + n_r H(f, g)$$

so that f^* is defined by the first order condition $-n_r \frac{\partial H(f,g)}{\partial f} = 1$.

If $\frac{\partial^2 H}{\partial f \partial g} = 0$, private protection has no effect on the government's choice of emergency defensive expenditures. If $\frac{\partial^2 H}{\partial f \partial g} > 0$, private investments g reduce the rate at which damages decrease with increases in f (the marginal benefit of emergency defensive expenditures), and thus the optimal choice of f during a disaster. For example, increased g may reduce a structure's vulnerability to wildfire, reducing the need for an aggressive firefighting response (the final possibility, $\frac{\partial^2 H}{\partial f \partial g} < 0$, seems unlikely in practice).

Knowing the central government's dispatch rule for aid during a disaster, homeowners in the risky place choose g to minimize their private disaster-related costs. When homeowners must reimburse the central government for their share of per-capita defensive expenditures, they solve

$$\min_g \quad g + \phi \frac{1}{n_r} f^*(n_r, g) + \phi H(f^*(n_r, g), g)$$

When homeowners do not pay for defensive expenditures, they solve

$$\min_{g} \quad g + \phi H(f^*(n_r, g), g)$$

The first order conditions for these problems are identical except for an additional $\frac{\phi}{n_r}\frac{\partial f^*(n_r,g)}{\partial g}$ term for the fully accountable household. This term is the marginal reduction in future expected emergency defensive expenditures due to investments in self-protection. Fully accountable households consider this benefit when choosing g. When emergency defensive expenditures are provided for free, households do not consider this benefit and thus choose less than the socially optimal investment in self-protection.

B Construction of the dataset

Our data combine administrative data on firefighting expenditures from multiple agencies, parcel-level assessor data for the universe of western US homes, topographical information, risk assessments, and weather conditions data. This section provides a complete account of the dataset construction; readers should refer to section Section 3 in the main paper for a high-level summary. Table 7 gives descriptive statistics for the dataset and Figure 8 maps all of the large fires in the sample, colored by agency.

B.1 Wildland Firefighting Expenditures

The fire suppression and preparedness cost data come from six different sources, including five federal agencies and one state firefighting agency. The federal agencies are the United States Forest Service, the National Park Service, the Bureau of Land Management, the Bureau of Indian Affairs, and the Federal Emergency Management Agency. The state agency is California's Department of Forestry and Fire Protection (Cal Fire). We obtained firefighting data at the incident level from each agency through a combination of Freedom of Information Act (FOIA) requests (or similar records requests for state data) and publicly available sources. Our geographical focus is the western United States. We define the "western United States" as the states of Arizona, California, Colorado, Idaho, Montana, New Mexico, Nevada, Oregon, Utah, Washington, and Wyoming. We discuss each source of data in detail below, as well as the process by which we harmonize these datasets.

B.1.1 US Forest Service

The US Department of Agriculture, Forest Service (USFS) accounts for the largest share of fire suppression expenditures of any federal agency and is primarily responsible for fires that ignite in or near the boundaries of National Forest areas. We obtain historical by-incident suppression costs (primarily wage and equipment costs incurred by USFS) for fires managed by the USDA Forest Service from 1995 to 2014 from the

National Fire and Aviation Management Web (FAMWEB) Database. Some institutional detail is helpful in understanding the process by which the data are compiled: the FAMWEB database represents a compilation of individual reports on fire occurrence, the conditions in which the fire ignited, and the suppression efforts undertaken by USFS. These reports are entered into the Fire Statistics System (FIRESTAT) application, which is run by the USFS. FAMWEB is the database which contains this information.[33]

Gebert, Calkin, and Yoder (2007) argue that fire suppression costs are captured more accurately by USFS accounting data than in the FAMWEB database. We therefore also obtain separate USFS accounting data on incident level expenditures through a separate Freedom of Information Act request. However, USFS was only able to provide these records for the period 2004–2012. Moreover, because of inconsistencies between agency reporting of incident PCodes, it is not possible to identify the fire characteristics for many fires in the accounting data. In Appendix Section C, we conduct our empirical analysis using both the accounting data and a subset of the FAMWEB data limited to 2004–2012 and find both qualitatively and quantitatively similar results. We conclude that inaccuracies in the FAMWEB database are sufficiently limited within our sampling frame to have limited impact on our empirical questions of interest and therefore conduct the bulk of our analysis with the FAMWEB data because of its greater temporal coverage.[34]

Over the course of our sampling frame, more than 150,000 wildfire incidents are logged in this database. However, since the Forest Service only reports per-fire cost data for fires above 300 acres, we limit this sample to the 2,419 fires in the 11 western states with a size of 300 acres or larger (the smallest size for which suppression expenditures are separately reported) for which the Forest Service was the jurisdictional owner. We also require that each fire have suppression cost, ignition date, and location data available.

Most ignitions are quickly suppressed at low marginal cost by "initial attack" efforts.

33. Previously, these data were compiled using Kansas City Fire Access Software, or KCFAST. Both KFCAST and FAMWEB include data on suppression expenditures and fire locations, but FAMWEB is the more current and complete of the two, with one exception: FAMWEB does not include any data on which agency was responsible for a given ignition or on the wind speed and direction at the nearest weather station at time of ignition. To obtain these additional fields, we also load and merge in the KCFAST dataset.

34. A more subtle difference between this study and Gebert, Calkin, and Yoder (2007) is that the latter authors use the fire cost per acre as the outcome variable when considering the drivers of wildfire suppression costs, arguing that "fire managers are accustomed to thinking in terms of cost per acre," and also include the natural log of total acres burned as an explanatory variable. We choose to use total cost as the outcome variable in our regression analysis of incident costs. We also do not include a measure of acres burned as an explanatory variable. We prefer this specification for two reasons: the policy-relevant figure is the total cost of suppression; and acreage burned as the denominator and size of fire as an explanatory variable induces a reverse causality problem (since acreage is a function of suppression effort) and a "bad controls" problem (Angrist and Pischke 2009).

These incidents are not included in our dataset of large fires. We address this in Section 5 by incorporating data on preparedness expenditures for USFS and the DOI agencies: these are expenditures that occur not in direct response to any particular large wildfire, but instead are undertaken to prevent or mitigate future fire risk. To identify these costs, we obtain budget justification reports from the US Forest Service website for the years 2007-2017. From these documents we extract the region-specific spending allocated towards "Fire Preparedness." In total we obtain more than \$6.1 billion of preparedness spending for the regions that overlap our sampling frame.[35] These preparedness costs represent the cost of maintaining initial attack readiness and other fixed costs of the wildland firefighting system. Section 5 describes how we allocate these costs over ignitions.

B.1.2 Department of Interior Agencies

Four separate agencies within the Department of Interior (DOI) engage in significant fire management. They are the Bureau of Land Management (BLM), the Bureau of Indian Affairs (BIA), the National Park Service (NPS), and the US Fish and Wildlife Service (FWS). We successfully obtained firefighting cost data for BLM, BIA, and NPS through FOIA requests. BLM is responsible for fires that ignite on the 248 million acres of public lands they manage. BIA is responsible for fires starting on the 55 million acres of Indian trust lands, and NPS is responsible for fires igniting within its 417 park units across 84 million acres of land. Each agency provided incident-level data from 2003-2016 from its own accounting databases for fires larger than 100–300 acres. To match the data available from the Forest Service, we limit this sample to include only fires that were the jurisdictional responsibility of the given agency and that affect more than 300 acres and apply similar data quality restrictions as those described for the USFS data. Our final DOI suppression dataset includes 1,617 BLM fires, 315 BIA fires, and 126 NPS fires.

As with USFS, we also include DOI preparedness costs in some scenarios in Section 5. The DOI agencies collectively prepare one annual budget justification that covers wildland fire activities across the entire United States. Our data on DOI preparedness costs come from the fiscal year 2012–2018 versions of these documents. In total, we account for \$2.7 billion of preparedness spending. Because DOI does not provide region-specific figures for these preparedness costs, we allocate them according to the proportion of total US ignitions that occur within our sampling frame on an annual basis. On average, we allocate 54% of this preparedness spending to our study area to obtain a total of \$1.5 billion from the DOI agencies.

B.1.3 California Department of Forestry and Fire Protection

We also collect fire suppression cost data for California, which includes over 50% of the population in our sample area and some of the most frequent and costly wildfires.

35. The Forest Service regions that overlap our sampling frame are 01, 02, 03, 04, 05, and 06.

Suppression cost data for California come from a public records request to the California Department of Forestry and Fire Protection (Cal Fire). Cal Fire is responsible for managing wildfires on 31 million acres of State Responsibility Area lands, loosely corresponding to private- and state-owned lands outside of incorporated towns and cities. We merge three sets of administrative records from Cal Fire. The first is a complete listing of all reported wildland fire incidents in the Cal Fire protection area during 2007–2016, regardless of size. This dataset includes the ignition date, acres burned, Cal Fire geographic unit, and, for incidents after mid-2011, the latitude and longitude of the ignition point.[36] The third dataset is an administrative record of firefighting expenditures at the incident level for 788 incidents during 2011–2016. According to Cal Fire, these expenditure data are carefully tracked because they are the basis of cross-agency reimbursements for mutual aid expenditures – for example, reimbursements to California by the federal government under the FEMA Fire Management Assistance Grant (FMAG) program, or by local governments to Cal Fire for firefighting assistance in incorporated areas.

Beginning with the list of significant fires, we drop those that are not the jurisdictional responsibility of Cal Fire. Limiting our sample to fires for which we are able to obtain precise location and suppression cost data results in 104 large fires (and 318 fires of any size) from 2011–2016.

B.1.4 Federal Emergency Management Agency

Our final agency source is the Federal Emergency Management Agency (FEMA). FEMA does not directly engage in firefighting efforts. Instead, FEMA reimburses state agencies and local governments for their costs on large firefighting efforts through the Fire Management Assistance Grant (FMAG) program. These grants reimburse 75% of the firefighting expenses incurred by state and local governments during qualifying incidents. We obtained incident-level data on FEMA reimbursements for wildfire incidents during 2000–2017 through a Freedom of Information Act request. These records contain the incident name, date, state, and amount reimbursed. They do not contain geographic coordinates (or a common identifier that would allow us to merge them to other agency data to recover geographic information). For cost scenarios in Section 5 that include FEMA reimbursements, we allocate these costs, multiplied by 1.33 to include the non-reimbursed portion, over fires in each year-state cell similarly to preparedness costs. In any calculation where we include Cal Fire cost data, we do not include FEMA reimbursements to California, which presumably include costs incurred by Cal Fire.

36. To supplement the location records for earlier fires, we also obtain shapefile data for a subset of Cal Fire incidents from the publicly available Fire and Resource Assessment Program database managed by Cal Fire.

B.1.5 Harmonization of Fire Suppression Cost Data

To ensure consistent data quality, we harmonize the data across all agencies from which we source suppression expenditures. Specifically, we ensure that ignition date, ignition location, responsible agency, cause of fire, area burned, and suppression cost data are present for all incidents and that the costs reflect values in 2017 dollars. Federal, state, and local firefighting agencies provide assistance to one another through coordinated dispatch systems and mutual aid agreements. We carefully considered the implications of this aid for our analysis. We confirmed with each agency that its reported costs represent only that agency's costs for a given incident (except for FEMA reimbursements). Thus, we avoid double counting when adding up historical costs across agencies in Section 5. When investigating the effect of homes on costs in Section 4.1, we use only USFS cost data and further limit the sample to incidents where USFS was the primary responsible agency. This restriction is used by Gebert, Calkin, and Yoder (2007), who argue that USFS bears at least 90% of the costs of these fires.[37]

We have also attempted to ensure that cost concepts are at least broadly comparable across agencies. In general, the firefighting cost data in the final dataset include wages (salaries, overtime, hazard pay) and equipment costs. Usage costs for agency-owned equipment (as opposed to equipment from private contractors) are tracked somewhat differently by different agencies. For example, in direct correspondence BLM indicated that they assign mileage costs for regular vehicles and engine-hour costs for fire engines to each incident, while NPS indicated that they assign only fuel and repair costs. The allocation of salary costs between "preparedness" and "suppression" budget categories may also differ somewhat across agencies.

Finally, we compute the spatial relationship between each fire and potentially valuable resources nearby. Specifically, we measure the distance from the ignition point of each fire to the nearest parcel in the parcels dataset described in Section B.2, the nearest state or federal highway, and the count of homes and their value within x km of the ignition point, where $x \in \{5, 10, \dots, 50\}$.

B.1.6 Ignition Point Characteristics and Weather Data

Using the harmonized location data, we obtain elevation, slope, aspect, and fuel model data for the ignition point of each fire from LANDFIRE. The former three products are derived from the high-resolution National Elevation Dataset; elevation represents the land height above sea level and is given in meters, slope represents the angle of the land and is given in degrees, and aspect represents the direction of the slope and is given in degrees as well. The fuel model data are the 13 Anderson Fire

37. Ideally, we would sum each agencies expenditures on each individual incident. Unfortunately, USFS and the DOI agencies do not reliably use consistent incident identifiers, making such a merge impossible.

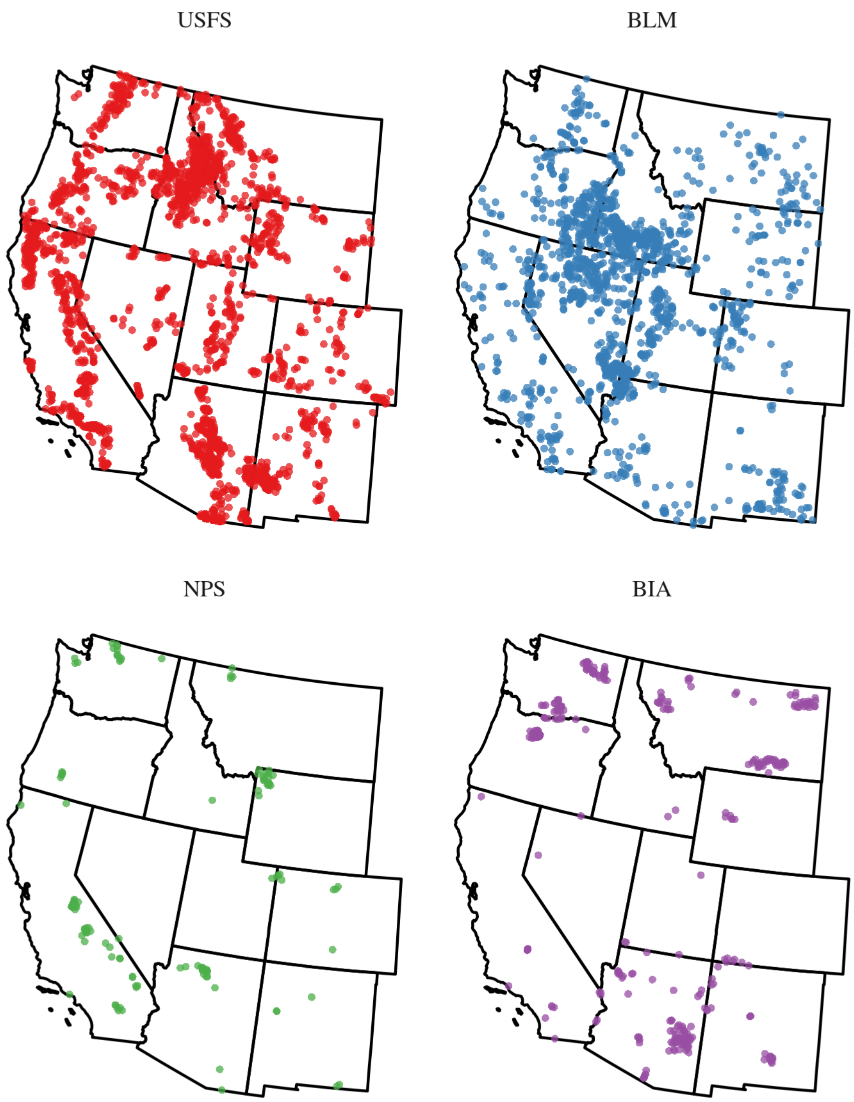
Appendix Table 7: Descriptive statistics

Panel A: Pooled fire characteristics	Mean	P10	P50	P90	
Area burned	7,873	383	1,433	16,034	
Fire cost	2,350,820	9,066	227,461	5,233,689	
Elevation	1,554	707	1,559	2,353	
Slope	12	1	10	29	
Temperature	20	13	21	27	
Precipitation	0	0	0	1	
Vapor Pressure Deficit	21	11	21	32	
Nearest home distance	14	1	10	33	
Parcels in 5km	160	0	0	109	
Parcels in 10km	757	0	0	1,011	
Parcels in 20km	3,345	0	90	7,093	
Value in 5km	45,633	0	0	18,536	
Value in 10km	210,261	0	0	182,871	
Value in 20km	936,119	0	13,094	1,450,741	
Panel B: Fire characteristics by agency	USFS	BLM	BIA	NPS	Cal Fire
Number of fires	2,419	1,617	315	126	104
Acres burned (1000s)	19,442	13,435	1,814	685	690
Suppression cost (m)	8,799	507	257	94	854

Notes: Table reports descriptive statistics for fires with area greater or equal to 300 acres in our sample. P10, P50, and P90 indicate the 10th, 50th (median), and 90th percentile of values. Aspect is given in degrees, elevation is in meters above sea level, fire cost is in 2017 US $, nearest home distance is in kilometers, parcels is the number of parcels within the given distance, precipitation is in mm, slope is in degrees, temperatures is in Celsius, and Vapor Pressure Deficit is in millibars.

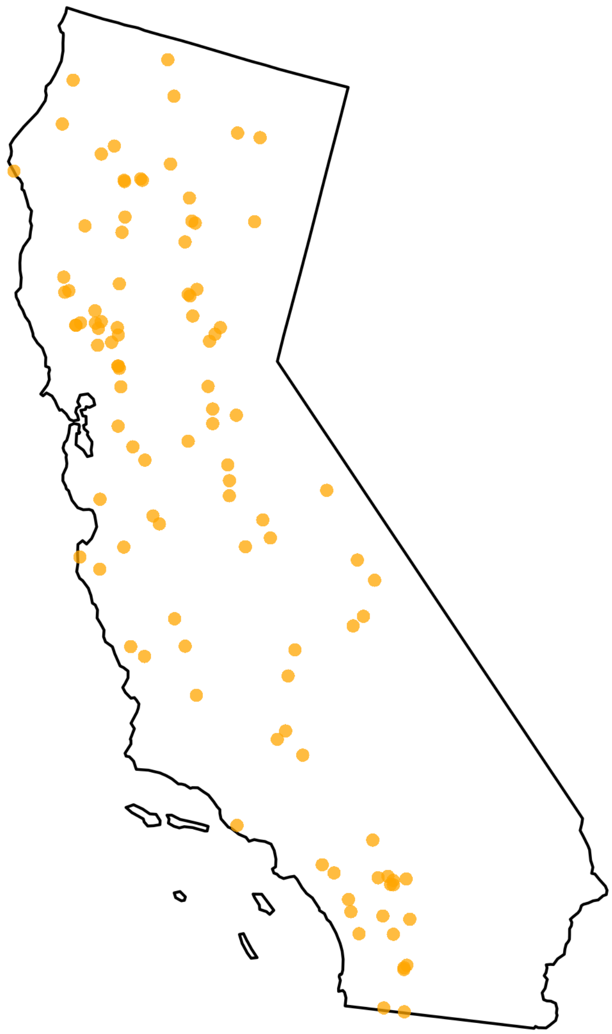
Behavior Fuel Models and describe the fire potential of surface fuel components (e.g., the type of foliage in the area). We also obtain ignition-day weather (maximum and minimum temperatures, precipitation, and measure of humidity) from the PRISM daily weather dataset, as well as ignition-day wind direction and speed from the FAMWEB dataset.

Appendix Figure 8: Federal Wildfires



Notes: Map of federally managed fires between 1995 and 2016 larger than 300 acres.

Appendix Figure 9: Cal Fire Wildfires



Notes: Map of Cal Fire-managed fires between 2011 and 2016 larger than 300 acres.

B.2 Parcel Data

The homes data include information on home locations, values, year built, and other property characteristics for 18.5 million parcels, or nearly all of the homes in the western United States. We also include parcels within 50 km of these states to accurately capture the nearness and number of parcels for wildfires that occur near the eastern borders of our sample. These data represent a compilation of tax assessor data from individual counties.[38] A primary advantage of these data is the inclusion of detailed locational information; specifically the data include both latitude and longitude as well as street address for each parcel. While previous studies in this area rely on publicly available data on the number and value of homes in a Census block (Gebert, Calkin, and Yoder 2007; Gude et al. 2013), this confidential dataset enables us to precisely locate homes relative to wildfire ignition points. Because Census blocks can be large in rural areas and particularly when located near national forests, the standard approach using Census block centroids introduces substantial noise into the estimate of distance-to-nearest parcel for each fire. In Section B.2.1 we document the improved locational precision and the data quality benefits produced by this approach.

We limit the sample to include only homes in partially vegetated areas that would be threatened by wildland fires, based on wildland-urban interface (WUI) categories identified in Radeloff et al. (2005). Specifically, we include homes located in the following vegetation categories: high density interface, high density intermix, medium density interface, medium density intermix, low density interface, low density intermix, very low density vegetated, and uninhabited vegetated.[39] We exclude homes in areas without wildland vegetation, and specifically in areas with the following categories: high density no vegetation, medium density no vegetation, low density no vegetation, very low density no vegetation, and uninhabited no vegetation. Because the federal government controls so much land in the West, and so much residential development is in wildland areas, these sample exclusions are not particularly restrictive. Our analysis dataset includes 9,148,972 homes (about 44% of all residential parcels including homes, condos, and apartments in the West).[40] We also link the parcels to the USFS Wildfire Hazard Potential (WHP) ratings to assess physical fire risk (Dillon 2015). These risk scores are designed to "depict the relative potential for wildfire that would be difficult for suppression resources to contain," and combine

38. This proprietary compilation was provided by CoreLogic© through a data agreement with Stanford University. Our comparisons to publicly-available home counts at the tract level, available upon request, confirm the comprehensiveness of the data.

39. Because the WUI data are built from Census records and our parcel data represent precise locations, occasionally a parcel is located in a so-called "uninhabited vegetated" area. As we rely on the WUI data to identify vegetated areas, we include homes in these areas as well.

40. This sample of 9.1 million homes used to estimate Equation (3) also includes homes near the sampling area but lying in bordering states in order to appropriately account for all nearby homes. In our main results, we report the expected protection cost only for homes in the 11 western states.

data from a large-scale fire simulator with spatial fuels and vegetation data to produce indicators of WHP. For each parcel, we assign a categorical and a continuous measure of WHP for that location as a measure of the risk faced by that parcel. We also add a measure of population density (population per square meter) from the Gridded Population of the World dataset, which reports density within roughly one km square grid cells.

The data also include reported transaction values. As is common for real estate data, many reported transactions do not represent true arms-length sales. We use only transaction values determined by CoreLogic to be arms-length transactions, and we further remove transactions indicated as refinancing, foreclosures, or inter-family transfers. We also exclude transaction values below $10,000 or above $100,000,000 in 2017 dollars, and transactions prior to 1980. After these cleaning steps, we have usable transaction values for 69% of homes in the raw data.

B.2.1 Comparison to Census Aggregate Data

Our study uses parcel-level data to assess the locations of homes threatened by wildfire. Previous studies rely on counts of housing units at the Census block scale (Gebert, Calkin, and Yoder 2007; Gude et al. 2013). Appendix Table 8 demonstrates that high-risk regions are systematically likely to have large Census block sizes. The average Census block size for homes in the highest decile of firefighting cost is 7.0 square km, and the 95th percentile is 29.7 square kilometers. This large grid size introduces substantial noise into geographic analyses of aggregate home counts. Our study instead uses parcel-level data to assess home locations. This represents a substantial increase in granularity over existing studies.[41] The degree of this advantage over aggregate block-level data depends on the accuracy with which parcel locations are reported in the real estate data. The underlying records in this dataset are collected by county tax assessors, and the quality of the data varies across counties. In the following section, we describe the process by which we obtain highly accurate parcel locations for the dataset and the advantages this provides relative to using Census block centroids.

The process of generating geographic coordinates for individual structure locations is called geocoding. This section compares the default geocoding for the homes in our dataset to an alternative geocoding algorithm. We also compare our results using methods to identify homes based on publicly available data that have been used in related work (e.g., Gebert, Calkin, and Yoder 2007; Radeloff et al. 2005; Radeloff et al. 2018).

The housing data used in this project come from a compilation of tax assessor data.

41. A separate advantage of parcel-level data over Census data is that we know the year in which a home was constructed, and thus whether the home was present at the time of each fire in the dataset. Census data report static housing counts every 10 years.

This dataset includes a field identifying the latitude and longitude of each home in the dataset. Overall, careful investigation of subsamples of the data imply that these coordinates are quite accurate. However, these default locations often locate multiple homes in precisely the same geographic location. To improve the accuracy of parcel locations, we implemented a secure, locally-hosted geocoding algorithm on a local server to calculate coordinates for each home. We used a locally hosted instance of the Nominatim geocoder[42] to geocode homes in our dataset based on the address field, while maintaining data confidentiality and security.

Overall, the geographic coordinates generated by Nominatim align closely with the default locations in the homes data. The median distance between reported locations is 41 meters. For most homes, we believe that the Nominatim locations represent small shifts that slightly improve location accuracy. The exception is for addresses that include typographical errors. In this case, Nominatim may return locations that are not meaningful – for example, that may be hundreds of kilometers outside of the county containing the home.[43] To eliminate these errors, we backstop the Nominatim locations with the default locations in the original dataset (which tend to be more accurate but less precise) using the following rule: if the Nominatim location is A) more than one km outside of the county given in the tax assessor data, B) differs from the tax assessor location by more than 5 km, or C) was not obtained using the street address (e.g., was geolocated by the Nominatim algorithm based only on city and state), we use the tax assessor location instead. Using this backstop method, we re-code 89% of the addresses in our full dataset using Nominatim, and the remainder with the default locations in the original dataset.

Previous studies of wildland-urban interface issues have used publicly-available Census data to identify approximate home locations. The decennial Census includes counts of population and housing units at the Census block level. Forestry studies frequently use these block-level aggregate data to locate homes (e.g., by average population over the area of the Census block, or assigning population to the centroid).[44] One challenge with using aggregate Census data is that Census blocks in areas with high fire risk tend to be many square kilometers or more, reducing the accuracy of the approach. Appendix Table 8 shows this. On the other hand, Census block-based approaches do not rely on the accuracy of address-based geocoding.

The figures and tables in this section explore the robustness of our results to three possible methods to locating homes: our geolocation method, a method that follows previous work in using Census block centroids for homes' locations, and a method

42. Nominatim uses Open Street Map data to conduct forward and reverse geocoding and is available at https://github.com/openstreetmap/Nominatim.

43. The County field in the underlying dataset is likely to be particularly reliable, since the dataset is assembled from individual county tax records.

44. Martinuzzi et al. (2015) describes one approach in detail, including how raw Census blocks are processed to remove portions that overlap public land and other steps.

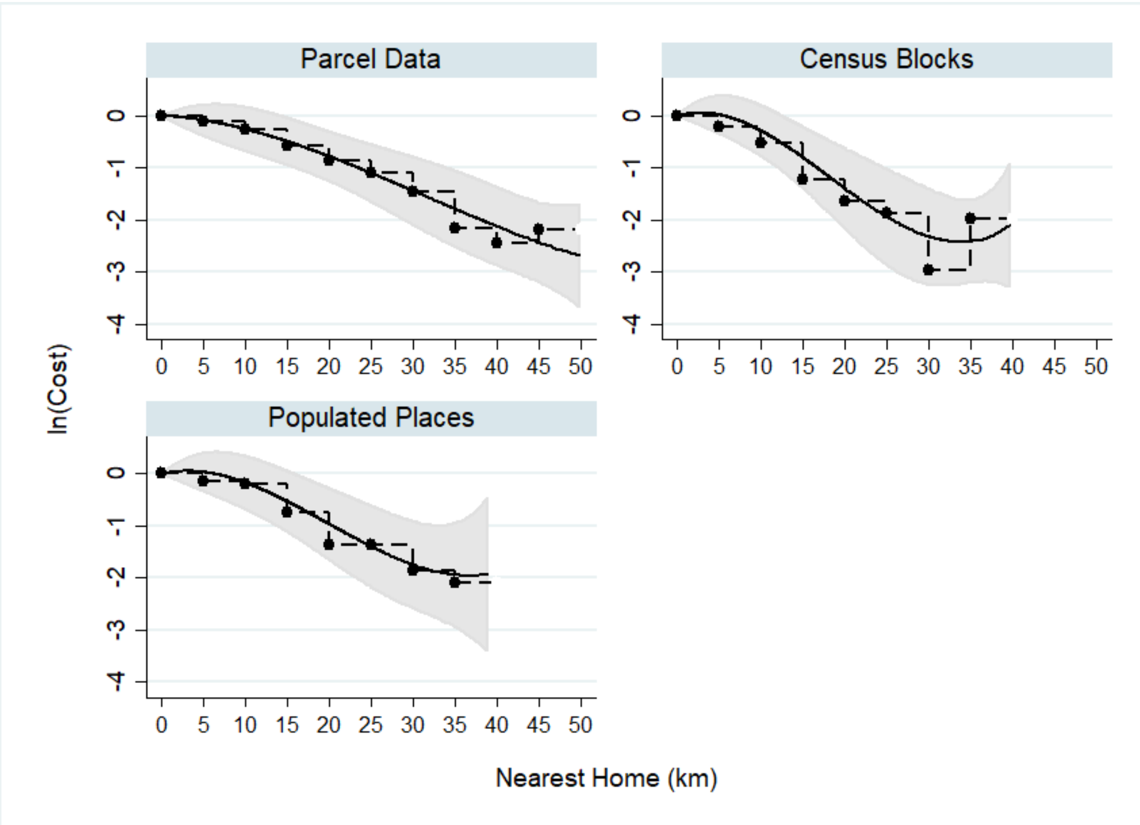
Appendix Table 8: The Advantage of Parcel-level Data: Census Blocks in High-Cost Areas are Large

	Area in km^2	
	All Populated Census Blocks	Highest Decile of Firefighting Cost
Mean	1.2	6.9
p90	0.9	14.7
p95	3.0	29.7
p99	22.8	101.9
N	416,983.0	42,021.0

Notes: This table shows the distribution of areas for Census blocks, in square kilometers. Column (1) includes all 2010 Census blocks with greater than zero housing units. Column (2) includes the 10% subset with the highest average expected protection costs as identified in our study. While Census blocks tend to be small overall, the areas of greater interest for understanding firefighting costs are systematically larger. Data on Census block areas, housing counts, and locations are from the US Census Bureau.

using the Census-based list of places (which include both incorporated and unincorporated communities). Appendix Figure 10 reproduces the regression from Figure 3 in the main text. The results are not qualitatively sensitive to the choice of location method. However, both of the Census-based approaches identify few fires with homes more than 40 km away and the corresponding standard errors for the estimate of the effect of home nearness on fire suppression cost are noisier. In our view, both of these facts reflect that the Census-based approaches systematically underestimate (on average) the distance to nearest home for fires in remote areas for the reasons we describe above.

Appendix Figure 10: Cost by distance to nearest home



Notes: Each panel estimates the impact of nearest home distance, as measured using three different methods of locating homes, on log suppression cost. "Parcel Data" uses the parcel real estate data with the geocoding and backstop method described in paper. "Census Blocks" uses Census block centroids. "Populated Places" uses the location information given in the Census Populated Places dataset. Each regression includes national forest fixed effects, state by month-of-year fixed effects, and state by year fixed effects. Standard errors are clustered by national forest.

B.3 Calculating Counterfactual Costs With No Nearby Homes

For each fire i, we use the regression results from Section 4 to calculate Δ_i, the increase in firefighting costs relative to what would have been spent on the incident if there were no nearby homes. This section describes that calculation and compares it to an alternative calculation based on a generalized linear model (GLM) approach.

B.3.1 Main Approach

Our main approach computes Δ_i using the binned model in Section 4.1. Consider a specification with 5 bins, corresponding to 0, 10, 20, 30, and 40+ kilometers distance to nearest home, where the omitted category is the 40+ kilometer bin. Let β_d represent the regression coefficient on the dummy variable for bin d. These coefficients give the increase in log firefighting costs when the nearest home is located d km away, relative to 40+ km. The percentage increase in firefighting costs in raw dollars can be calculated as $e^{\beta_d - 0.5s} - 1$, where s is the sample analog of the variance of β_d (Halvorsen and Palmquist 1980; Kennedy 1981). In other words, the regression provides an estimate of the average effect of distance to nearest home on firefighting costs. We use these average effect estimates to calculate counterfactual costs in the absence of any homes within 40 km. For homes in bin d, letting c_i be the observed cost and $\tilde{c}_i$ the counterfactual cost, we calculate $\tilde{c}_i = \frac{c_i}{e^{\beta_d - 0.5s}}$. Then Δ_i is $c_i - \tilde{c}_i$.

B.3.2 Alternative Approaches: GLM and Retransformation

These counterfactual costs could be computed in other ways. A similar approach with the same OLS semi-log regression is to use the regression coefficients to generate predicted log costs under the counterfactual, and then "re-transform" these predicted values to predictions in dollar units (Duan 1983; Manning et al. 1987; Manning 1998). These counterfactual predicted costs can then be subtracted from predicted costs given the observed distance to home, $\hat{c}_i$. In practice, the various retransformation estimators are vulnerable to specification error, especially in the presence of heteroskedasticity (Manning and Mullahy 2001).

A potentially more attractive approach is to use a statistical model that does not require retransformation. Instead of semilog OLS, Manning and Mullahy (2001) recommends the use of a generalized linear model (GLM) with a log link function. Among other advantages, the GLM model generates predicted values in raw dollar units. We implement the GLM approach as a check on the robustness of our main estimates. Following the results of the selection algorithm in Manning and Mullahy (2001), we use a GLM model with a gamma distribution and a log link.[45] With the GLM approach, Δ_i can be calculated either by using the implied average change in costs in each distance bin (as we did for the OLS estimates), or by directly generating predicted costs given the observed and counterfactual x's. We show results

45. See page 471 in Manning and Mullahy (2001). The resulting value of λ is about 2.3.

for both approaches. Table 9 shows that the average predicted cost differences are similar across approaches. The approach using OLS generates slightly smaller predicted cost differences, implying that the cost differences we use in the main text are conservative.

Appendix Table 9: Counterfactual cost differences

Observed distance	(1) OLS	(2) GLM	(3) GLM
Panel A. Average Percentage Change in Costs			
0-10	86	88	88
10-20	80	86	86
20-30	66	77	77
30-40	30	45	45
40+	0	0	0
Panel B. Average Dollar Difference (thousands)			
0-10	4,113	4,207	4,662
10-20	2,874	3,070	3,170
20-30	1,351	1,573	1,669
30-40	397	599	299
40+	0	0	0

Notes: Panel A shows the average percentage decrease in cost for an otherwise-identical fire with no homes within 40 km. Panel B shows the average difference in expenditures for an otherwise-identical fire with no homes within 40 km (in thousands of dollars). Column (1) uses the percentage changes implied by the semilog OLS regression coefficients to scale the observed costs. Column (2) uses the percentage changes implied by the GLM regression coefficients to scale the observed costs. Column (3) also uses GLM, but reports the difference in predicted costs using the observed values of the covariates and predicted costs with no homes within 40 km.

C Comparison to Forest Service Accounting Data

Our main analysis makes use of publicly available data on suppression expenditures for US Forest Service Fires. However, Gebert, Calkin, and Yoder (2007) write that the publicly available data on costs are less accurate than official expenditure data recorded in the USFS accounting system. Since the time of their writing, the addition of an accounting code (known as a “P-code”) to the FAMWEB data has made this match somewhat more straightforward.

To check whether the results of our empirical exercise in Section 4.1 are altered by the use of the more accurate accounting data, we submitted a Freedom of Information Act Request to the US Forest Service for the accounting dataset. The dataset we obtained as a result of this processing includes suppression expenditures from 2003-2013 with a limited set of fields. Specifically, it includes the P-code, the amount of suppression expenditures for that code, and the year that those expenditures were billed. The following table summarizes yearly cost for 2004-2012 (2003 and 2013 are partially missing in the accounting dataset) for the FAMWEB data and the accounting dataset we obtain.

Appendix Table 10: Annual costs by suppression cost dataset

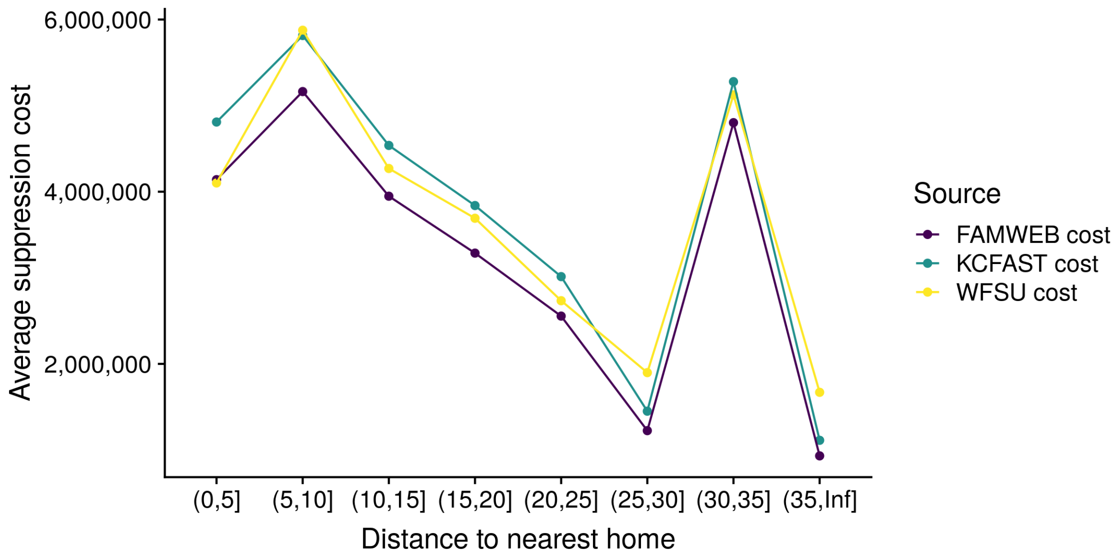
Year	FAMWEB	FAMWEB West	WFSU valid	WFSU all
2004	247	236	471	679
2005	271	262	440	768
2006	828	799	1,142	1,355
2007	978	923	977	1,263
2008	708	694	1,070	1,464
2009	401	394	682	840
2010	239	224	373	662
2011	475	436	623	1,251
2012	975	952	917	1,161
Total	5,122	4,920	6,695	9,442

Notes: All values in millions of dollars. First column includes all incidents in FAMWEB, second column includes only incidents in regions 01-06, third column includes only WFSU incidents with P-codes used for wildfire suppression-related costs. Specifically, the incident code begins with P*, where * is a number for the USFS region, and is followed by a 4 character alphanumeric code beginning with a letter, per USFS specification.

Next, we match the costs in the accounting dataset to the FAMWEB data using the P-code to identify whether the relationship between suppression costs and distance from homes is stable across the use of either source of cost data. We match from the P-code and year to the suppression expenditure data from FAMWEB. This match is not entirely straightforward: the guidelines over the issuance of P-codes and the proper accounting procedures have changed over the years, and many fires are submitted under the same P-code. In particular, large complex fires are often accounted for using the same P-code.[46] For the 997 fires in our FAMWEB dataset from 2004-2012,

46. So-called "ABCD" fires, which are small, are also accounted for using a single P-code for each forest-year, but for our purposes this is not an issue since our focus is on incidents with more than 300 burned acres.

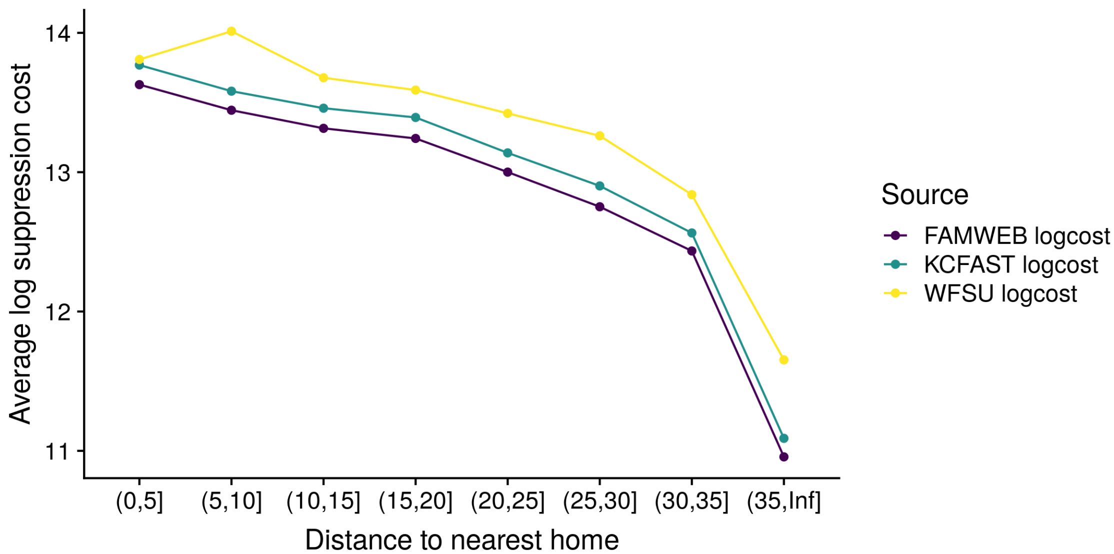
Appendix Figure 11: Comparison of FAMWEB and accounting data: mean suppression costs and distance to nearest home



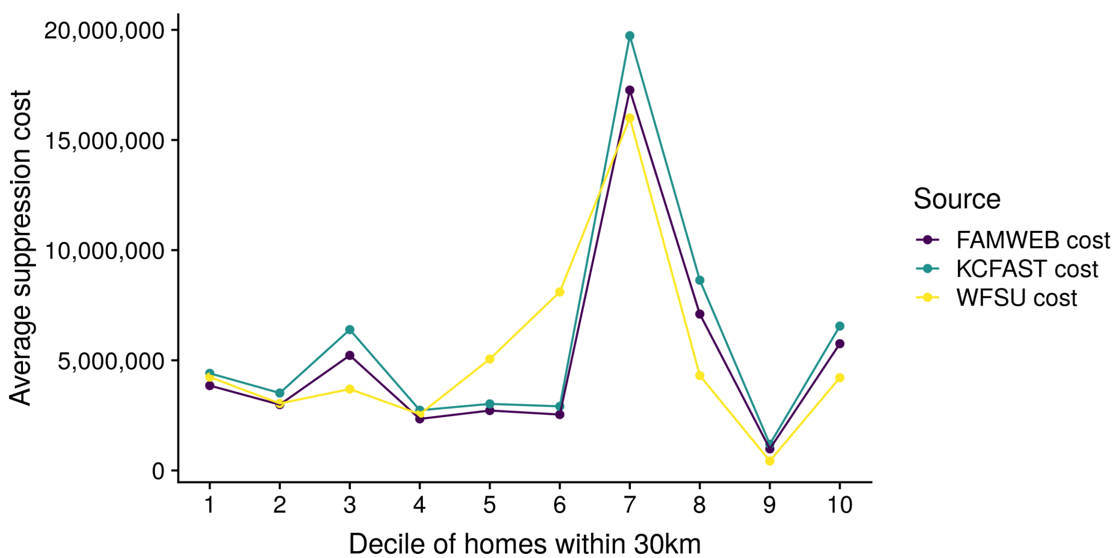
we are able to match 799 of these to the accounting dataset.

We estimate the relationship between fire cost and nearby homes for four sets of costs: A) FAMWEB costs for all fires in FAMWEB, B) FAMWEB costs for all 2004-2012 fires in FAMWEB, C) FAMWEB costs for fires that match to the accounting data, and D) accounting data costs for all fires that match to FAMWEB data. Figures 11 to 14 plot binned averages and sums of costs for each dataset on distance from nearest home and on number of homes within 30km. Although the sums differ due to the difference in the number of fires included for each set of data, the means have similar patterns. Our conclusions about the usefulness of the FAMWEB data are similar to those of Schuster, Cleaves, and Bell (1997), who wrote at the time that, “One of the purposes for our analysis of per-acre fire expenditures was to assess the quality of suppression expenditure estimates contained in the NIFMID database. These estimates are widely regarded as unreliable. However, the correlation between uncorrected, NIFMID-based expenditures and those from the accounting system is 0.85, a surprisingly high level.”

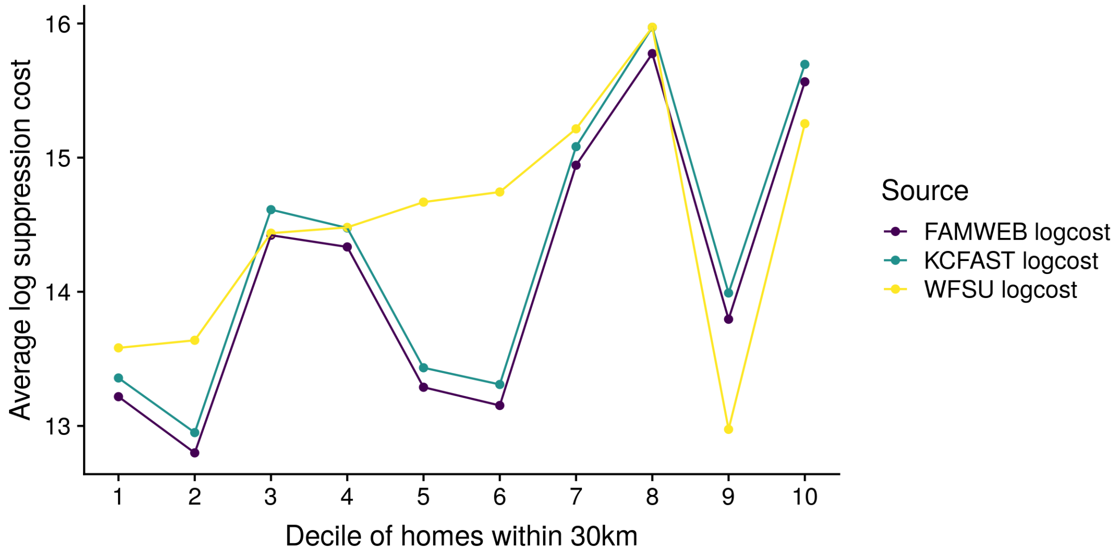
Appendix Figure 12: Comparison of FAMWEB and accounting data: mean log suppression costs and distance to nearest home



Appendix Figure 13: Comparison of FAMWEB and accounting data: mean suppression costs and number of homes in 30km



Appendix Figure 14: Comparison of FAMWEB and accounting data: mean log suppression costs and number of homes in 30km



Appendix References

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling." *The Journal of Human Resources* 40 (4): 791–821.

Angrist, Joshua, and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* 1st ed. Princeton University Press.

CIESIN. 2017. *Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 10.* Palisades, NY.

Dillon, Gregory K. 2015. *Wildfire Hazard Potential for the Coterminous United States.* Forest Service Research Data Archive.

Duan, Naihua. 1983. "Smearing Estimate: A Nonparametric Retransformation Method." *Journal of the American Statistical Association* 78 (383): 605–610.

Gebert, Krista M, David E Calkin, and Jonathan Yoder. 2007. "Estimating Suppression Expenditures for Individual Large Wildland Fires." *Western Journal of Applied Forestry* 22 (3): 188–196.

Gude, Patricia H., Kingsford Jones, Ray Rasker, and Mark C. Greenwood. 2013. "Evidence for the Effect of Homes on Wildfire Suppression Costs." *International Journal of Wildland Fire* 22.

Halvorsen, Robert, and Raymond Palmquist. 1980. "The Interpretation of Dummy Variables in Semilogarithmic Equations." *American Economic Review* 70 (3): 474–75.

Kennedy, Peter. 1981. "Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations." *American Economic Review* 71 (4): 801.

Manning, Willard G. 1998. "The Logged Dependent Variable, Heteroscedasticity, and the Retransformation Problem." *Journal of Health Economics* 17 (3): 283–295.

Manning, Willard G, and John Mullahy. 2001. "Estimating Log Models: To Transform or Not to Transform?" *Journal of Health Economics* 20 (4): 461–494.

Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, and Arleen Leibowitz. 1987. "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment." *The American Economic Review* 77 (3): 251–277.

Martinuzzi, Sebastián, Susan Stewart, David Helmers, Miranda Mockrin, Roger Hammer, and Volker Radeloff. 2015. *The 2010 Wildland–Urban Interface of the Conterminous United States.* Research Map NRS-8. U.S. Department of Agriculture, Forest Service, Northern Research Station.

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* 37 (2): 187–204.

Radeloff, Volker, Roger Hammer, Susan Stewart, Jeremy Fried, Sherry Holcomb, and Jason McKeefry. 2005. "The Wildland–Urban Interface in the United States." *Ecological Applications* 15 (3): 799–805.

Radeloff, Volker, David Helmers, H. Anu Kramer, Miranda Mockrin, Patricia Alexandre, Avi Bar-Massada, Van Butsic, et al. 2018. "Rapid Growth of the US Wildland-Urban Interface Raises Wildfire Risk." *Proceedings of the National Academy of Sciences* 115, no. 13 (March): 3314–3319.

Schuster, Ervin, David Cleaves, and Enoch Bell. 1997. *Analysis of USDA Forest Service Fire-Related Expenditures 1970—1995.* Research Paper PSW-RP-230. USDA Forest Service Pacific Southwest Research Station, March.

USDA (USDA Office of Inspector General). 2006. *Forest Service Large Fire Suppression Costs.* Audit Report 08601-44-SF.