Equity and Adaptation to Wildfire Risk: Evidence from California Public Safety Power Shutoffs

Max Vilgalys*

August 25, 2022

Abstract

In the past decade, California investor-owned electric utilities have begun implementing Public Safety Power Shutoffs (PSPS) as part of their effort to adapt to increasing risk of catastrophic wildfires. I examine the extent that these decisions are correlated with two measures of community vulnerability: health risk factors and socioeconomic status (SES). I first construct a dataset linking weather, vulnerability indices, and PSPS decisions for electric circuits in California's three largest investor-owned utilities. I show that PSPS is used more frequently in circuits with lower average SES among two of California's major utilities, and circuits with higher average health risk in one of the major utilities. To focus on utilities' decisions, rather than other sources of inequality that may place vulnerable communities in areas with higher wildfire risk, I repeat this analysis after controlling for weather variation. The results are qualitatively similar. I then model the utility's decision problem, as reported in regulatory filings, and measure which components of the model may be responsible for the PSPS decisions. After controlling for weather variation, I find that ignitions are more frequent in low-SES circuits and in lower health risk circuits for one utility. I cannot reject that utilities' estimated costs from declaring PSPS shutoffs or expected damages from wildfires are equitably distributed.

1 Introduction

In the last decade, electric utilities in California have been forced to adapt to increasing risk of catastrophic wildfire. Climate change, forest management practices, and shifting wildland-urban

*`vilgalys@mit.edu` Massachusetts Institute of Technology, Institute for Data, Systems, and Society.

interface have contributed to the most severe wildfire seasons in California's history. Some of the costliest wildfires have been sparked by California electric utilities, and utilities are responsible for those financial damages under California law. Utilities have already faced billions in dollars in fines, notably driving Pacific Gas and Electric (PG&E) to declare bankruptcy in 2019. To make their electric lines safer, utilities invest in managing vegetation, upgrading infrastructure, and moving lines underground. However, these improvements are relatively slow and wildfire risk can require utilities to respond quickly.

This paper focuses on utilities' last-ditch effort to prevent wildfire, the Public Safety Power Shutoff (PSPS). In a PSPS, a utility preemptively de-energizes lines that are likely to spark large wildfires. PSPS imposes concentrated costs to impacted communities, and diffuse benefits to the utility and the overall public if utilities are successfully preventing wildfire. These shutoffs are subject to strict regulation, and utilities must demonstrate that they carefully weigh the costs and benefits of each de-energization event (CPUC, 2019). The first PSPS was in 2013, and since then over 5,000 circuits (small segments of the electric grid) have been de-energized and over 1 million customers impacted (Hill et al., 2020). I focus on shutoffs by the three largest investor-owned utilities in California: Southern California Edison (SCE), Pacific Gas and Electric (PG&E), and San Diego Gas and Electric (SDG&E). Utilities must disclose PSPS events, unlike other investments for wildfire resilience that are protected as information for critical national infrastructure. This makes PSPS decisions a useful data source to examine the equitability of utilities' adaptation to wildfire risk.

While PSPS is necessary in the short run to adapt to rising wildfire risk, shutoffs could exacerbate inequalities if disadvantaged communities receive fewer benefits or bear more costs. This paper focuses on evaluating the distribution of costs of PSPS. It is challenging to evaluate the benefit of PSPS because this depends on predicting the damages from wildfires that might occur without these shutoffs. Predicting wildfire size and damages are notoriously challenging problems, even given modern machine learning techniques (Taylor et al., 2013; Xi et al., 2019; Jain et al., 2020). Utilities use proprietary models and datasets to predict wildfire size, and only declare a PSPS event when they believe that there is a high probability of catastrophic wildfire (SCE, 2021; PG&E, 2021; SDG&E, 2021a). Many Californians benefit from reduced wildfire risk, and rural populations or those with health risk factors may benefit most from reducing wildfire smoke (D'Evelyn et al., 2022). I do not attempt to model the benefits of PSPS, but note that these shutoffs are likely preventing catastrophic wildfires and their associated health risks.

I examine how the costs of PSPS are distributed by measuring the extent that PSPS decisions have been equitably targeted. Equitable targeting means that communities with the same observable risk factors (e.g. weather variation), but differ in health risk factors or socioeconomic status (SES), experience the same rate of shutoffs. I focus on health risk and SES as measures of vulnerability because low-SES communities may have limited resources to adapt to electricity failures and those with health risks may experience health complications from wildfire smoke or electricity outages. I use definitions of health risk and SES from CalEnviroScreen (August et al., 2021). Correlation between PSPS decisions and vulnerability indices could be explained by differences in treatment by the utilities or by various weather, vegetation, or development conditions that impact wildfire risk. If this correlation persists after accounting for all factors outside the utility's control, actions by the utility (such as unequal infrastructure investments or biased PSPS decision rules) are responsible. I compare circuits that differ in average SES or health risk, with and without population and weather controls. This provides evidence of whether vulnerable populations experience different rates of shutoffs and whether that is explained by observable factors outside the utility's control.

To conduct this analysis, I first construct an extensive dataset linking vulnerability indices, weather, and PSPS records from 2014-2021. For vulnerability, I use Census tract-level data on health risk factors and SES indices from August et al. (2021). The health risk index includes rates of asthma, cardiovascular disease, and low birth weights. The SES index includes unemployment, rates of high school completion, and linguistic isolation. My weather observations come from GridMET, a dataset of daily observations of 13 weather variables used to predict wildfire size (Abatzoglou, 2013). I use records of PSPS outages and ignitions along power lines from filings to the California Public Utilities Commission (CPUC). I merge all these datasets with geospatial records of circuits from utilities' Integrated Capacity Analysis maps.

I find that there are significant associations between vulnerability scores and PSPS decisions, although these vary by utility. Without controlling for weather factors, circuits with lower SES are significantly more likely to have a shutoff in PG&E and SDG&E (at the $p = 0.001$ level), and less likely in SCE (p value 0.005). Populations with higher health risk are significantly more likely to have a shutoff in SDG&E (at the $p = 0.001$ level) and less likely (at the $p = 0.001$ level) in PG&E. After controlling for population and weather variation, model fit improves but these patterns remain largely consistent. This shows that the difference in rates of PSPS by vulnerability indices is largely unexplained by population or weather differences.

To better understand what factors lead to these decisions, I develop a model of PSPS decisions based on guidelines in utilities' published Wildfire Mitigation Plans (SCE, 2021; PG&E, 2021; SDG&E, 2021a). When weather conditions suggest that large fires are likely, utilities identify circuits that could spark a large wildfire. Teams of meteorologists, fire scientists, and data scientists predict regions where ignitions are likely to spread to large fires. They use data of line conditions from public weather reports, service crews, and private weather stations to identify lines that could spark wildfires. These experts form predictions using these rich sources of public and private information, relying on machine learning models and extensive simulations of wildfire behavior. If their predictions find that the likely costs of wildfire exceed the costs of shutting off power, they notify residents and de-energize the circuit. Power remains off until weather conditions are less severe and the utility inspects affected circuits for any debris or damage.

I examine how health risk and SES indices correlate with the probability of an ignition and the firm's cost of declaring PSPS. To find the probability of ignition, I use logistic regression with records of ignitions along circuits from 2014-2021. I find that ignitions are significantly (at the $p = 0.001$ level) more frequent in low-SES circuits and in lower health risk circuits in PG&E, after controlling for weather variation. To find the cost firms incur from declaring PSPS, I examine Wildfire Mitigation Plans and post-event reports. In their regulatory filings, utilities state that the cost of declaring PSPS is linear in the expected size of interruption (SCE, 2021; PG&E, 2021; SDG&E, 2021a). However, the calculations I find from post-event reports indicate that cost is linear in number of customers (Valdberg, Tozer, and Kilberg, 2021). The results using either metric as a proxy are fairly noisy, and I do not observe a strong association between vulnerability indices and the utility's cost of declaring PSPS.

I can infer how vulnerability indices correlate with the firm's estimated cost of wildfire by imposing a model of the utility's decision problem. With my model, the estimated coefficients relating probability of PSPS, probability of ignition, and utility's cost of declaring PSPS imply the coefficients on the missing component: the utility's estimated damage from a wildfire. I do not access this data directly, because utilities use complex, proprietary software to project wildfire damages. After controlling for weather and population, my results are fairly noisy. The only significant finding (after controlling for population and weather variation, at the $p = 0.001$ level) is that PG&E estimates higher wildfire damages in lower-SES circuits. Overall, I cannot reject that utilities are equitably estimating damages.

This project relates to several literatures. First is a literature studying the environmental justice of wildfire risk. In early work to explore this topic, Niemi and Lee (2001) describe how poverty can increase wildfire incidence and damages and Ojerio (2008) shows that federal wildfire preparedness grants are concentrated in higher-SES communities. One strand of this literature focuses on comparing populations that live in high wildfire risk regions. Wigtil et al. (2016) document that places with higher wildfire potential generally have lower social vulnerability to wildfire risk. Wibbenmeyer and Robertson (2022) find higher average property value, older residents, and more white residents in places with high wildfire potential. Another strand focuses on the impacts and responses of wildfires. D'Evelyn et al. (2022) argue that the health effects of wildfire smoke disproportionately impact populations with limited adaptive capacity. Anderson, A. Plantinga, Wibbenmeyer, et al. (2020) study inequality in firefighting responses, and document preferential treatment to higher SES communities following salient wildfire events. A. J. Plantinga, Walsh, and Wibbenmeyer (2022) study the historical spread of fires and find that firefighting efforts prioritize high-value properties.

Within this literature, several recent studies have examined PSPS as a tool to combat wildfire risk. Guliasi (2021) gives an analysis of the political economy and history of the PSPS. Hill et al. (2020) examines potential health costs from PSPS, and Wong-Parodi (2020) surveys impacted California residents about attitudes towards PSPS events. Rhodes, Ntaimo, and Roald (2020) studies the PSPS as an optimization problem, and suggests improvements to current decision processes using a test case. My paper is the first, to my knowledge, to empirically study the equity of these shutoff decisions.

This project is also related to a literature on identifying bias in decision making, specifically in cases where agents make decisions relying on complex algorithms. There is a broad literature on studying discrimination in decision-making, dating back to at least Becker (1957). Lang and Kahn-Lang Spitzer (2020) and Mehrabi et al. (2021) provide reviews of economics and machine learning literature, respectively, on identifying bias in decision making. Recent examples examining bias in human decisions include an analysis of racial bias in healthcare decision rules (Obermeyer et al., 2019) and in pretrial appearance risk (Rambachan, 2021). Examples examining bias in algorithms include facial recognition software (Buolamwini and Gebru, 2018) and predicting risks from medical records data (Gianfrancesco et al., 2018; Parikh, Teeple, and Navathe, 2019). Like these studies, I examine decisions and look for evidence of unequal treatment after controlling for relevant, exogenous variation. I focus on a setting that is less well-studied in the literature, where agents make algorithm-supported decisions.

My project is also related to literature on measuring equity in adaptation to climate change. Among environmental advocates, there has long been a call to focus on equity in climate change adaptation (Smit and Pilifosova, 2003; Thomas and Twyman, 2005). In their report, IPCC (2022) identifies several settings where inequality and poverty have set "soft limits" on the ability of groups to adapt to climate change. Coggins et al. (2021) conducted a review of literature on equity in climate change adaptation and highlighted several examples of work assessing the equity of climate adaptation. Sheller and Leon (2016) use interviews to study how historical inequalities between Haiti and the Dominican Republic impacted government responses to similar environmental crises, and Satyal, Byskov, and Hyams (2021) use environmental justice theory to examine how systemic injustices facing an indigenous group in Uganda undermine adaptation planning. However, Coggins et al. (2021) ultimately conclude that more work is needed in this area, especially in empirical assessment of equity and justice. This paper addresses this by providing more work on empirical assessment of equity and justice in these shutoff decisions.

The remainder of this paper is structured as follows. Section 2 describes the various datasets used in the analysis, and provides summary statistics. Section 3 describes my modeling approach including references to utilities' filings that justify my modeling decisions. Section 4 gives the results of my analysis, and discusses their interpretation. Section 5 concludes.

2 Data

I use a variety of sources to construct a dataset with circuit-level records of weather variation, vulnerability indices, and shutoff decisions from 2014-2021. The unit of observation is a circuit-day. An electric circuit is a small unit of the electricity distribution network, and generally the level at

Utility	year	2013	2014	2017	2018	2019	2020	2021
SCE	Customers	–	–	–	–	196,879	235,879	117,690
	Million CMI	–	–	–	–	353	280	372
	# PSPS Events	–	–	–	–	246	1,501	122
PG&E	Customers	–	–	–	47,324	1,987,783	645,859	79,630
	Million CMI	–	–	–	89.8	6,670	1,560	174
	# PSPS Events	–	–	–	32	1,458	670	219
SDG&E	Customers	179	884	17,111	21,036	45,337	93,058	–
	Million CMI	0.0797	0.665	40.5	65.4	78.2	165	–
	# PSPS Events	3	6	51	38	218	110	–

Table 1: Number of PSPS events by firm, by year, and the number of customers impacted. CMI is Customer Minutes Impacted, the product of the minutes of shutoff and number of customers per circuit. Note that number of customers impacted is the sum of customer shutoffs experiences, but not the unique number of customers impacted.

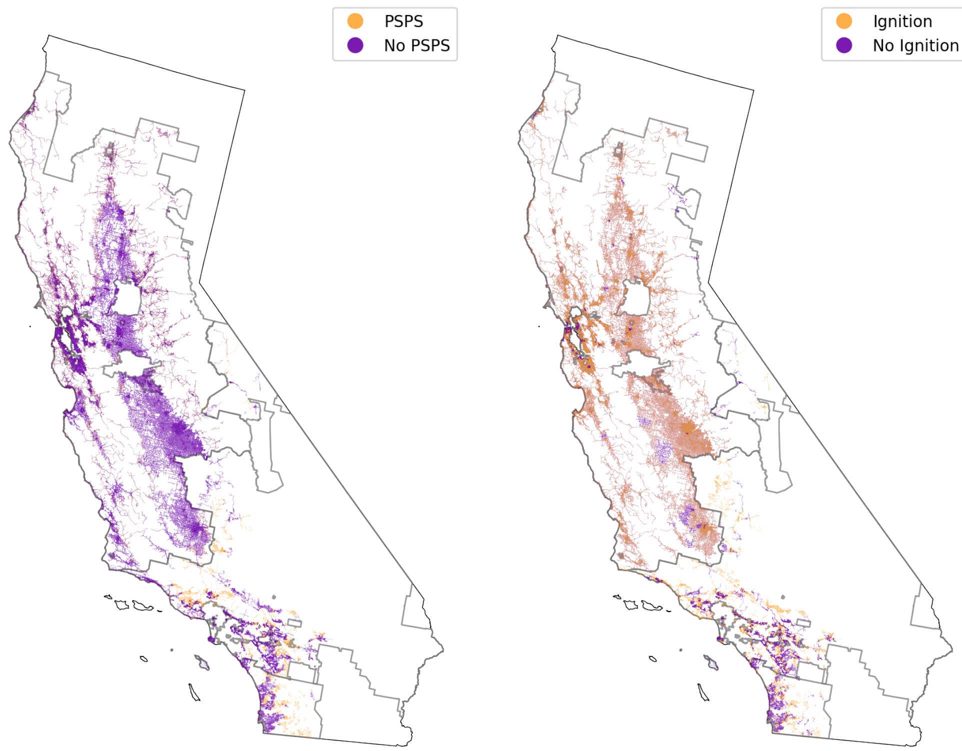
which PSPS decisions are recorded. I treat vulnerability as fixed over the sample period.

2.1 PSPS Events

For PSPS events, I use filings from firms to the CPUC. Firms are required to report statistics after each shutoff, so this dataset represents the universe of shutoffs between October 2013 and December 2021. The CPUC summarizes these reports and publishes a record of each shutoff. Each record includes the circuit targeted, the date and time of the shutoff, the duration of the outage, the number of customers impacted, and information on what types of customers are impacted. Table 1 summarizes these filings by year and firm.

PSPS events are generally reported at the circuit level. In some cases, a firm reports a sub-circuit level outage. I sum these outages to the circuit level to match the weather records in my data.

In order to link these with other geospatial records, I use integrated capacity analysis (ICA) maps from each electric utility. ICA maps are circuit-level maps of the distribution infrastructure, although some circuit segments are not published due to privacy concerns. I am able to match over 98% of PSPS records to their corresponding geographic file. The ICA maps include 5,411 circuits; there are PSPS events recorded on 20.3% of these circuits. Figure 1a shows the location of these circuits.



(a) PSPS Events

(b) Ignitions

Figure 1: Map of which circuits had PSPS events (a) and ignitions (b) from my data. State borders and boundaries between electric utilities are shown.

2.2 Fire Data

My main analysis uses data on fire ignitions along utility lines from filings to the CPUC from 2014 through 2021. Per CPUC guidelines, firms must report all fires to their knowledge larger than one meter (CPUC, 2014). This dataset includes 4,550 ignitions from the three firms I study. These filings are required to include the ignition location, but not the corresponding circuit segment. To match these to the circuit records, I find the closest circuit segment from the ICA maps to the ignition location. Figure 1b shows the location of these circuits.

2.3 Vulnerability

I use indices from CalEnviroScreen to measure population vulnerability (August et al., 2021). The authors construct a Census tract-level database of health risk factors and socioeconomic status (SES) indicators. I use these indexes, as well as the tract-level population, in my analysis. This database is primarily intended to assess environmental and energy justice in the state of California.

In the main analysis, I summarize these data with an SES index and a health risk index. Each index ranges from 0-100, with 100 being the most vulnerable and 0 being the least. The indices are constructed as the average of ranks of several factors, as in August et al. (2021). For socioeconomic vulnerability, this includes rate of high school non-attainment, rent-burdened low-income households, limited English proficiency, living below twice the federal poverty line, and share unemployed. For the health risk index, this includes asthma incidence, cardiovascular disease incidence, and rate of low birth weight infants.

To match these records to circuits, I take the average of values from each census tract that contains a given circuit segment. I weight these averages by the length of the circuit in each census tract. I am able to match records for 5,000 out of 5,411 circuits, and for 1,071 of the 1,103 circuits with a PSPS event.

Figure 2 plots these scores per circuit against the total number of PSPS events (among circuits with at least one event), the total number of recorded ignitions (among circuits with at least one ignition), and the total customer minutes interrupted (among circuits with at least one event). Each plot also includes the best-fitting line to these observations, to help summarize the trend among these scatter plots.

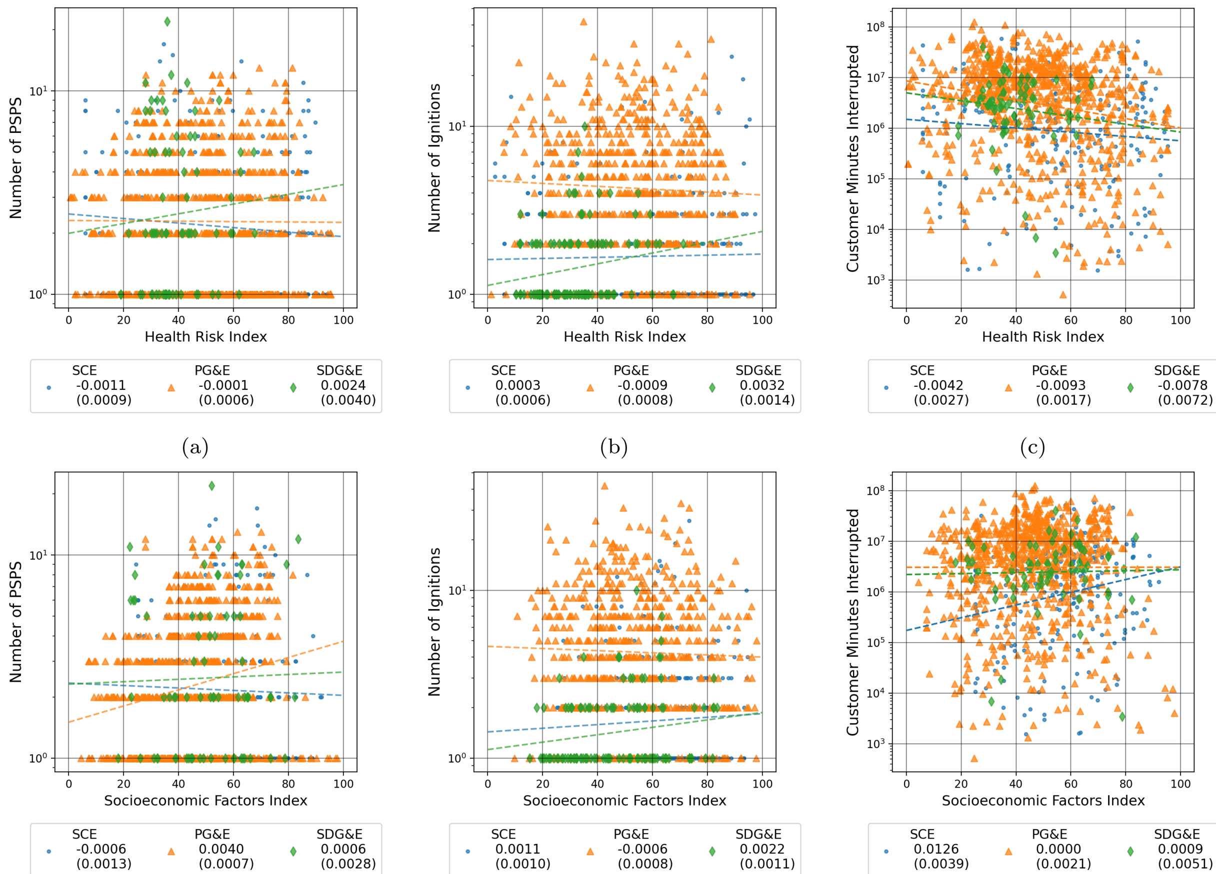
2.4 Weather data

For weather observations, I use the GridMET weather dataset from Abatzoglou (2013) and an archive of areas with a red flag warning. This dataset was designed to support applications in modeling wildfire risk, and includes a rich set of relevant weather variables. Each variable is reported daily at a high spatial resolution (4 km) across the United States; I include observations from California.

GridMET includes primary variables, constructed via satellite- and geography-guided interpolation from weather stations, and variables derived from these primary observations. To merge GridMET records with my dataset, I find all grid points within $2\sqrt{2}$ km of a circuit and take the simple average of weather records at each observation. Primary variables are specific humidity, precipitation, minimum relative humidity, maximum relative humidity, surface downwelling shortwave flux in air (a measure of solar radiation), minimum air temperature, maximum air temperature, wind speed, and wind direction. Derived variables are expected to be relevant for predicting wildfire

	(1)	(2)	(3)
Max Air Temperature (C)	23.80	28.17	23.31
	(8.048)	(6.364)	(6.390)
Min Air Temperature (C)	10.01	11.91	9.897
	(5.564)	(5.073)	(4.850)
Precipitation Amount (daily mm)	1.265	0.0595	0.00312
	(5.598)	(0.775)	(0.0579)
Specific Humidity (kg/kg)	0.00639	0.00521	0.00394
	(0.00216)	(0.00251)	(0.00167)
Wind Velocity at 10 m (m/s)	3.318	3.670	5.325
	(1.581)	(1.655)	(2.102)
Wind From Direction (Degrees past North)	233.6	228.1	224.2
	(82.76)	(90.89)	(111.6)
Mean Vapor Pressure Deficit (kPa)	1.274	1.945	1.565
	(0.946)	(0.836)	(0.694)
Max Relatively Humidity (%)	77.82	57.08	50.94
	(19.37)	(21.26)	(19.13)
Min Relatively Humidity (%)	33.80	16.62	15.40
	(18.47)	(11.80)	(9.803)
Surface Downwelling Shortwave Flux (W/m^2)	223.8	232.1	190.1
	(96.88)	(74.95)	(47.95)
Burning Index (Derived)	36.15	54.96	68.55
	(20.91)	(16.57)	(19.45)
Energy Release Component (Derived)	46.07	67.08	70.08
	(23.77)	(14.89)	(14.03)
Potential Evapotranspiration (Derived, mm)	4.204	5.440	5.092
	(2.409)	(2.138)	(1.783)
Reference Evapotranspiration (Derived, mm)	5.769	7.994	7.978
	(3.342)	(3.076)	(2.787)
Dead Fuel Moisture 100 hr (Derived, %)	12.91	8.538	7.821
	(5.100)	(3.055)	(2.591)
Dead Fuel Moisture 1000 hr (Derived, %)	14.03	9.771	9.301
	(5.494)	(2.631)	(2.689)
Elevation	241.0	314.8	493.6
	(325.2)	(382.9)	(400.1)
Observations	16303343	652310	3333

Table 2: Summary statistics of GridMET data from October 2013 through 2021. Columns separate the full sample, the sample during a red flag warning, and the sample during a PSPS event. Observations are weighted by the length of each circuit segment. Standard errors of the mean for each column are in parentheses.



(a) (b) (c)

(d) (e) (f)

Figure 2: Scatter plots showing vulnerability indices and various outcomes, for circuits with nonzero values of PSPS events or ignitions. Each plot includes a best-fitting line for the observations. The coefficient on the index and and the standard error (in parentheses) of each line are reported in the legend.

risk: burning index, energy release component, potential evapotranspiration, reference evapotranspiration, dead fuel moisture at 100 hours, and dead fuel moisture at 1000 hours. See Abatzoglou (2013) for more details on the development of this dataset.

In addition to weather variation, firms use Red Flag Warnings from the National Weather Service to make PSPS decisions. Over 98% of shutoffs occur during a Red Flag Warning, a period when the National Weather Service has identified weather conditions that could sustain catastrophic fires. I include an indicator of whether a Red Flag Warning was in effect in any part of a circuit by merging a historical archive of Red Flag Warning shapefiles.[1]

Summary statistics of each weather variable are given in Table 2. Relative to the full sample, Red Flag Warnings are drier, hotter, and more elevated, and PSPS events occur in windier and

[1] From `https://mesonet.agron.iastate.edu/info/datasets/vtec.html`, accessed 9 December 2021.

drier conditions and in higher locations.

3 Model

My empirical model comes from firms' descriptions of their shutoff decision process. Per their filings to the CPUC, firms initiate a PSPS if the expected degree of damages (that is, the product of expected damages conditional on ignition and the probability of ignition) exceeds the cost of failing to provide power.[2] I am interested in the degree that health risk factors and SES are correlated with PSPS decisions, and whether this is due to the channels of expected damages from wildfire, probability of ignition, or the firm's cost of declaring a PSPS event. I am unable to directly estimate the degree that these vulnerability indices are associated with expected damages, but I can infer this parameter from my model and estimates of the two other channels.

The shutoff decision is a binary choice model, where the firm weighs the expected damages from a wildfire ("Wildfire Risk") against the firm's cost of failing to provide power ("PSPS Risk"). Per filings to the state regulator, firms use separate prediction problems for probability of an ignition and size of fire conditional on ignition (PG&E, 2021; SCE, 2021; SDG&E, 2021a). This approach is common in both classical statistical (Xi et al., 2019) and machine learning (Jain et al., 2020) approaches to predicting wildfire size. This implies that fire size and ignition probability are conditionally independent, given an appropriate set of controls. Conditional independence allows environmental factors and income to influence both fire damages and ignition probability, but assumes that other shocks to fire ignition are unrelated to fire damages. This allows firms to have private information of shocks that influence either ignition probability or fire size (e.g. having a line crew detect a fallen tree along a power line), as long as those shocks provide no additional information on the other outcome variable conditional on our control variables.

I therefore write the firm's problem as:

$$\text{PSPS}_i = \mathbf{1}\left\{\text{Prob.}(\text{ignition}_i|X_i, Z_i)\mathbb{E}[C(\text{damages}_i)|X_i, Z_i] \geq \mathbb{E}[\{\text{PSPS Risk}\}_i]\right\} \tag{1}$$

where X_i is the set of vulnerability indices of interest, and Z_i are additional controls including weather variation, elevation, and population.

I use a generic function C for the expected degree of damages because firms' profit function is likely not linear in damages. If firms face nonlinearly increasing consequences from large fires (e.g.

[2] Legislation requires that firms making shutoff decisions must quantify benefits and risks of de-energization events, and document "how the power disruptions to customers, residents, and the general public is weighed against the benefits of a proactive de-energization."

bankruptcy, as PG&E experienced after the 2019 fire season) or are risk averse.[3]

Let $\pi(X_i, Z_i) := \mathbb{E}[\mathbf{1}\{\text{ignition}_i\}|X_i, Z_i]$ be the conditional probability of ignition, $\phi(X_i, Z_i) := \mathbb{E}[U(\text{damages}_i)|X_i, Z_i]$ be the firm's expected cost from damages conditional on ignition, and $\mathbf{V}(X_i, Z_i) := \mathbb{E}[\{\text{PSPS Risk}\}_i|X_i, Z_i]$ be the firm's expected cost from declaring PSPS. The firm faces uncertainty over the cost of PSPS because this depends on the duration of an outage, which is determined by how long weather factors remain in effect. Replacing each term by its expectation and taking logs, I can rewrite Equation (1) as a sum of these expectations plus an expectational error term:

$$\text{PSPS}_i = \mathbf{1}\left\{\log(\pi(X_i, Z_i)) + \log(\phi(X_i, Z_i)) - \log(\mathbf{V}(X_i, Z_i)) + \epsilon_i \geq 0\right\} \tag{2}$$

where ϵ_i is an expectational error term.

To summarize how factors X_i influence decisions, I introduce a partially linear approximation to each function $\pi, \phi, \mathbf{V}$.

$$\log \pi(X_i, Z_i) = \sigma_1 X_i + \tilde{\pi}(Z_i); \quad \log \phi(X_i, Z_i) = \sigma_2 X_i + \tilde{\phi}(Z_i); \quad \log \mathbf{V}(X_i, Z_i) = \sigma_3 X_i + \tilde{\mathbf{V}}(Z_i) \tag{3}$$

Let γ be the overall nuisance function: $\gamma(Z_i) := \log(\tilde{\pi}(Z_i)) + \log(\tilde{\phi}(Z_i)) - \log(\tilde{\mathbf{V}}(Z_i))$. Then, I can estimate the following equation:

$$\text{PSPS}_i = \mathbf{1}\left\{(\sigma_1 + \sigma_2 - \sigma_3)X_i + \gamma(Z_i) + \epsilon_i \geq 0\right\} \tag{4}$$

I make the standard assumption that ϵ_i is a type-I extreme value random variable, and estimate this decision as a logistic model. The coefficient on X_i from this logistic regression is the overall degree that vulnerability indices influence PSPS decisions, $(\sigma_1 + \sigma_2 - \sigma_3)$. I assume that γ is a linear function of log population and weather variables, although it is also possible to use a more flexible approach. By taking hypothesis tests on whether this overall parameter is different from zero, I evaluate the research question of whether circuits with different vulnerability indices experience different rates of shutoffs.

I also wish to find which parts of the firm's decision problem explain any differences in the rates of shutoffs between circuits with different vulnerability indices. In terms of the model, I wish to test whether each parameter $\sigma_1, \sigma_2,$ and σ_3 is significantly different from zero. I am able to estimate σ_2 (the contribution to the overall coefficient from ignition probabilities) and σ_3 (the contribution to the overall coefficient from PSPS cost) through separate regression problems, but am not able

[3] PG&E explicitly includes nonlinear risk weighting in their decision function; see PG&E (2021), section 4.2.a, for a description, and justifies this behavior as risk aversion in PG&E (2020).

to estimate σ_1 directly.

The value of σ_1 is implied given estimates of σ_2, σ_3, and $(\sigma_1 + \sigma_2 - \sigma_3)$. I cannot estimate σ_1 directly, as I am not able to estimate the damages from a fire or the firms' cost functions based on those damages. Damages from a wildfire are a function of wildfire size and the features of land damaged by the wildfire. Predicting fire size is a notoriously challenging problem, even given modern machine learning techniques (Taylor et al., 2013; Xi et al., 2019; Jain et al., 2020). Firms use proprietary software to make these fire size predictions, and achieve relatively high levels of accuracy. In Appendix A, I document an attempt to predict fire size using linear regression and random forest regressions. I am unable to provide informative bounds on the degree of fire size, and am therefore unable to estimate the expected damages or the firm's expected damages (after applying the cost function).

To estimate σ_2, I use a two-stage procedure with data of ignitions along utility lines. To estimate σ_3, I use a regression of proxies to the firm's cost, according to their filings. I discuss these estimation problems in subsections below.

3.1 Ignition Probability

I estimate σ_2 (the contribution to the overall coefficient from ignition probabilities) by first finding the probability of ignition, and then finding the coefficient on vulnerability indices X_i from a partially linear model to those predicted probabilities. To estimate the probability of ignition, I use logistic regression of ignitions along power lines with control variables $\{X_i, Z_i\}$. I adjust the standard errors from the second stage estimates, as the second stage estimates depend on the results from the first stage.

To construct the probability, I model ignition probability as a binomial logit model. I assume that there exists some latent model of fire ignitions, for some function g of vulnerability indices and environmental factors and a type-I extreme value distributed error term ε_i:

$$\text{ignition}_i = \mathbf{1}\{g(X_i, Z_i) + \varepsilon_i > 0\} \tag{5}$$

With this model, I can calculate the ignition probability $\hat{\pi}(X_i, Z_i)$ given an estimate of g:

$$\hat{\pi}(X_i, Z_i) = \frac{\exp \hat{g}(X_i, Z_i)}{1 + \exp \hat{g}(X_i, Z_i)} \tag{6}$$

I estimate g using a subset of data from years where firms do not use PSPS. Table 1 shows the years with PSPS observations. I use data from all firms in 2015 and 2016 and from a subset of firms in 2014, 2017, 2018, and 2021. When firms use PSPS, data is censored: the researcher does not observe whether an ignition would have occurred without PSPS. With endogenously censored outcome variables, it is generally only possible to partially identify regression functions (Khan and Tamer, 2009). Subsets from years when firms do not use PSPS are not subject to this censoring concern. I assume that the relationship between weather variation and ignition probability is consistent between years when utilities do and do not use PSPS; this could be violated if dry vegetation accumulates and fire risk increases over time, or utilities choose to use other wildfire management strategies in years without PSPS. In Appendix C, I conduct a robustness exercise and estimate ignition probabilities using the full sample. While I do not fully characterize the partially identified set, these exercises support my main findings.

To estimate σ_2, I regress the log probability computed from the first stage on X_i and Z_i. Per Equation (3), I assume that the log probability is additively separable in X_i and Z_i. The parameter σ_2 can then be estimated via partially linear regression.

$$\log \hat{\pi}(X_i, Z_i) = \sigma_2 X_i + \tilde{\pi}(Z_i) + \varepsilon_i \tag{7}$$

As this second stage depends on the first-stage estimation, I adjust standard errors for σ_2 to incorporate uncertainty in my estimate of g. I do so by finding the influence function of the first-stage problem and incorporating this into standard errors of the second-stage problem, as in Newey and Daniel McFadden (1994). Appendix B shows the derivation.

I implement these stages to estimate σ_2. I use linear models for the functions g and $\tilde{\pi}$. The procedure is as follows:

1. Estimate $\hat{g}$ via logistic regression of ignitions on X_i and Z_i. With a linear model, estimating $\hat{g}$ means finding a parameter vector $\beta = \{\beta_0, \beta_1, \beta_2\}$ that maximizes the likelihood of the model $\text{ignition}_i = \mathbf{1}\{\beta_0 + \beta_1 X_i + \beta_2 Z_i + \varepsilon_i > 0\}$.

2. Find $\hat{\sigma}_2$ using Equation (7) and the results from the first stage. For the log probabilities, use the estimated first-stage parameter vector to compute the probabilities $\hat{\pi}(X_i, Z_i)$. Then find σ_2 via linear regression of these log probabilities on X_i and Z_i. To find standard errors for σ_2, I adjust the standard errors from linear regression to incorporate error in estimating $\hat{g}$.

3.2 Cost of PSPS

I estimate σ_3 (the contribution to the overall coefficient from PSPS cost) through regression of proxies to the firm's cost of PSPS, based on their filings to regulators. Due to ambiguity between various documents, I use both the number of customers impacted and the customer minutes interrupted (CMI) as proxies to the firm's cost of PSPS. These proxies capture major sources of variance in the firm's expected costs of PSPS. While neither is a perfect approximation, they provide a reasonable estimate of how vulnerability indices influence the firm's cost of PSPS.

In firms' Wildfire Mitigation Plans, they state formulas to calculate cost of a PSPS that depend on customer minutes interrupted (CMI) and the total number of customers interrupted (SCE 2021, p. 61; PG&E 2021, p. 52; SDG&E 2021a, p. 26). Firms also incorporate the safety cost and financial cost of PSPS, as well as a reliability score. This safety cost is calculated as a constant factor multiplied by CMI, and the financial cost scales with the cost of shutoff (SCE, 2022; PG&E, 2020; SDG&E, 2021b). PG&E incorporates a scaling function if the safety, reliability, or financial costs of PSPS in a circuit exceed 10% of the largest recorded wildfire damages; I assume that the damage at any circuit never exceeds this threshold. SDG&E plans to incorporate the health sensitivity of subpopulations, but I do not observe decisions made with these rules (SDG&E, 2021a, p. 30). In 2021, SCE began weighting some components of its cost function by the number of vulnerable customers per line; I do not have access to their conversion formula and do not attempt to model this improvement. While the conversion factors are not published, this is absorbed into the constant if I take a regression of log CMI or log number of customers.

SCE is the only firm to specify how they form ex-ante predictions of the CMI. In their post-event reports, SCE calculates their CMI as a constant number of minutes multiplied by number of customers impacted, effectively making the cost of a shutoff a function of function solely of the number of customers (Valdberg, Tozer, and Kilberg, 2021, p. 16). No other firms publish their ex-ante PSPS cost calculations. I assume that they either use a constant factor, or the expected CMI per outage based on the empirical duration of PSPS outages.

With these proxies selected, I estimate σ_3 using linear regression on the log of the proxy. The number of customers impacted and CMI for a given outage are stochastic; depending on weather conditions, firms may be able to de-energize a smaller section of the circuit or be forced to prolong

	(1)	(2)	(3)	(4)	(5)	(6)
Health Risk Index	53.28	50.01	50.16	48.21	39.65	39.19
	(1.610)	(1.160)	(1.010)	(0.439)	(1.130)	(0.768)
SES Index	54.18	57.34	49.01	48.65	50.35	51.91
	(1.149)	(0.712)	(0.758)	(0.349)	(2.018)	(1.193)
Observations	223	469	451	1885	81	181
Utility	SCE	SCE	PGE	PGE	SDGE	SDGE
$\geq$24 Hours		X		X		X

mean coefficients; se in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Summary of vulnerability indices (SES index and health risk index), by whether the observed outage exceeds 24 hours. Standard error of the mean is in parentheses.

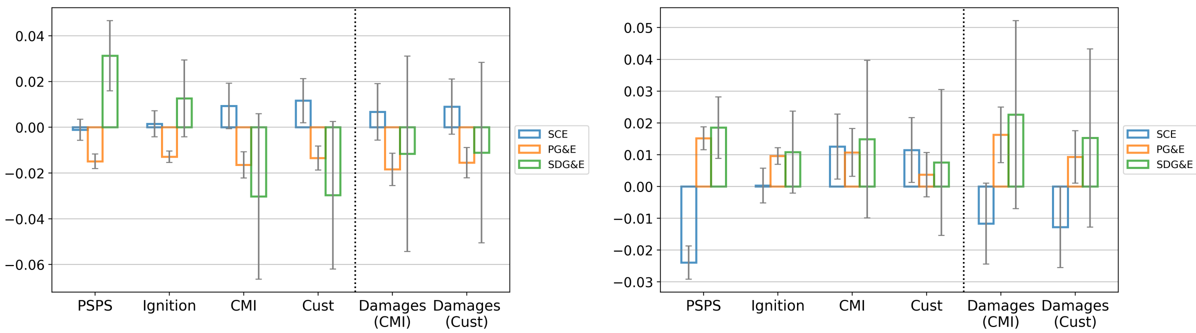
the outage. I estimate this relationship using outage duration and number of customers impacted during each reported outage, with vulnerability indices, weather controls, and log population as additional controls. I assume that for circuits with zero reported shutoffs, there is the same relationship between average vulnerability indices and CMI.

If firms use a constant outage duration to estimate costs, this approximation may systematically undervalue the cost to low SES or high health risk communities. To inspect this, I compare the average health risk and SES indices for circuits with PSPS outages above and below 24 hours. Table 3 shows these summary statistics. For SCE, outages over 24 hours occur in circuits with significantly higher SES index (indicating lower-SES circuits), and significantly lower health risk index. This shows that SCE's stated decision systematically undervalues the cost of an outage to low-SES populations.

4 Results

As each firm has unique decision rules, I report separate coefficients for each firm in all trials. All regressions include utility-by-year fixed effects, and all population or weather control variables are interacted with these fixed effects. In each subsection, I use four specifications: no controls (beyond fixed effects), only population as a control variable, primary weather variables plus population, and all weather variables plus population.

The dependent variables of interest are the health risk index and socioeconomic factor index from CalEnviroScreen. Recall that in these indices, 0 is the least vulnerable and 100 is the most vulnerable. Increasing the socioeconomic or health risk index by 1 is equivalent to an average



(a) Health Risk Index

(b) SES Index

Figure 3: Collected coefficient estimates of health risk or SES index on various outcomes, using all weather and population controls. Each group of plots on the x axis show coefficient estimates for a given outcome variable. Damages (to the right of dotted line) are implied from estimates to the left of dotted line and my model. Error bars show ± 1.96 times the standard error. Each group of plots is ordered SCE, PG&E, SDG&E.

increase of 1 percentage point across the ranks of the sub-indices. A positive coefficient indicates that the event in a given logistic regression is more likely, or the expected outcome in a linear regression is larger, when the population on the circuit has higher average health risk or lower average SES. I refer to circuits where the population has a lower (higher) average health risk index as lower (higher) health risk circuits, and circuits where the population has a lower (higher) average SES index as lower (higher) SES circuits.

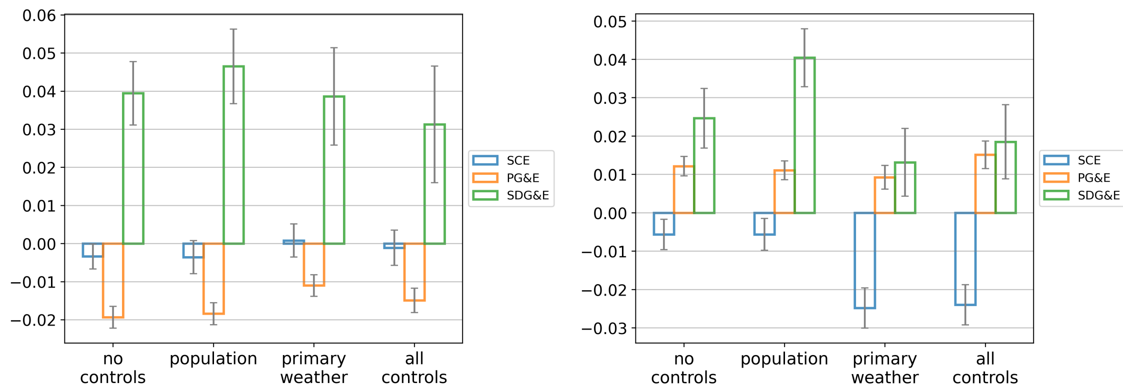
Figure 3 summarizes the overall coefficient estimates, for regressions including population and all weather controls. This figure shows that, after controlling for my set of covariates, there is a significant association between PSPS decisions and my vulnerability characteristics, in some utilities. The estimates of CMI and number of customers impacted (and the damage estimates that depend on these values) are especially noisy and I am unable to reject that these metrics are equitably distributed. Additional summary figures are shown in Appendix D using different sets of weather variation.

The following subsections describe the results from each separate regression.

4.1 PSPS Decisions

To study PSPS decisions, I use a subset of data during red flag warnings, from October 2013 (the month of the first PSPS event) onward. I limit the scope to red flag events to recreate the firm's problem, as over 90% of PSPS events are declared during a red flag warning. Coefficient estimates from the logistic regression are reported in Table 4, and visualized in Figure 4.

A positive coefficient indicates that higher health risk (lower SES) circuits have a greater rate



(a) Health Risk Index

(b) SES Index

Figure 4: Coefficient estimates of vulnerability indices for logistic regression of PSPS events. Groups on the x axis collect results from regression with a given set of controls. Error bars show ±1.96 times the standard error. Each group of plots is ordered SCE, PG&E, SDG&E.

of PSPS shutoffs. These coefficients represent the amount that log-odds change with an increase of 1 unit of the index. They can be approximately interpreted as the percentage change in PSPS likelihood given an increase in 1 unit of the index, as the coefficients are fairly close to 0. For example, an estimated coefficient of 0.01 indicates that PSPS events are 1% more likely in circuits with 1 higher index.

Without controlling for weather factors (column 1), higher health risk circuits are significantly less likely to have a PSPS in both SCE and PG&E, and more likely in SDG&E. This finding is significant at the $p = 0.001$ level for PG&E and SDG&E, and at the $p = 0.05$ level for SCE. Lower SES circuits are more likely to have a shutoff in PG&E and SDG&E, and less likely in SCE; this finding is significant at the $p = 0.001$ level for PG&E and SDG&E, and at the $p = 0.01$ level for SCE. These magnitudes are on the order of 0.01, so a 1 point increase in the index corresponds to roughly one percent difference in the likelihood of PSPS.

After controlling for weather variation (columns 2-3), model fit improves but these patterns remain largely consistent. D. McFadden (1973) suggests that a pseudo-R squared of 0.2-0.4 suggests good model fit for logistic regression, indicating that our model acceptably fits the PSPS decisions after controlling for weather variation. The exception is that the coefficient on the health risk index for SCE is no longer significant, but the coefficient on socioeconomic factors for SCE is larger in magnitude and is statistically significant at the $p = 0.001$ level.

I do not observe the full set of relevant variation that firms have while making these decisions.

	(1)	(2)	(3)	(4)
	psps	psps	psps	psps
SCE x Health	-0.00335*	-0.00355	0.000820	-0.00109
	(0.00167)	(0.00222)	(0.00222)	(0.00236)
SCE x SES	-0.00565**	-0.00562**	-0.0248***	-0.0240***
	(0.00202)	(0.00211)	(0.00268)	(0.00267)
PG&E x Health	-0.0193***	-0.0184***	-0.0110***	-0.0149***
	(0.00145)	(0.00147)	(0.00143)	(0.00163)
PG&E x SES	0.0122***	0.0111***	0.00926***	0.0151***
	(0.00129)	(0.00124)	(0.00157)	(0.00183)
SDG&E x Health	0.0394***	0.0465***	0.0386***	0.0313***
	(0.00425)	(0.00498)	(0.00652)	(0.00781)
SDG&E x SES	0.0247***	0.0404***	0.0131**	0.0185***
	(0.00397)	(0.00385)	(0.00451)	(0.00494)
Observations	375064	375064	370263	370263
Pseudo R^2	0.081	0.090	0.297	0.378
Population		X	X	X
Primary			X	X
Derived				X

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Results from logistic regression of PSPS events. Perfectly predicted failures are omitted. A positive coefficient indicates that higher health risk (lower SES) circuits have a greater rate of PSPS shutoffs.

In linear models, Oster (2019) gives an approach to quantify the degree of omitted variable bias by comparing the stability of coefficients as the model fit improves. I am not aware of an analogous approach for logistic regression. Informally, the sign and magnitude of coefficients remain relatively stable as the model fit improves from the null model to the model including population and all weather controls, indicating that these conclusions may be robust to incorporating additional variables.

4.2 Ignition Probability

I measure how vulnerability indices influence ignition probability via a two-stage procedure, as described in Section 3.1. In the first stage, I use logistic regression of ignitions versus vulnerability indices, log population, and weather covariates. In the second stage, I find the best fitting linear approximation to the log probability given the first-stage results. Standard errors from the second stage are adjusted to account for uncertainty from the first stage estimation, as described in Appendix B.

As described in Section 3.1, I use data from years where utilities did not conduct PSPS. This

	(1) ignition	(2) ignition	(3) ignition	(4) ignition	(5) ignition	(6) ignition	(7) ignition	(8) ignition	(9) ignition
Health Risk Index	-0.00314	-0.0113***	0.0213**	-0.00178	-0.0111***	0.0162*	0.00652*	-0.0121***	0.0113
	(0.00209)	(0.00114)	(0.00665)	(0.00264)	(0.00115)	(0.00723)	(0.00305)	(0.00127)	(0.00859)
SES Index	0.00308	0.0148***	0.0174**	0.00244	0.0146***	0.0196**	-0.00404	0.00916***	0.00955
	(0.00258)	(0.00110)	(0.00664)	(0.00262)	(0.00110)	(0.00625)	(0.00278)	(0.00133)	(0.00666)
Observations	1950168	4939641	700344	1950168	4939641	700344	1950168	4939641	700344
Pseudo R^2	0.006	0.007	0.013	0.007	0.007	0.017	0.047	0.060	0.080
Utility	SCE	PG&E	SDG&E	SCE	PG&E	SDG&E	SCE	PG&E	SDG&E
Population				X	X	X	X	X	X
Primary							X	X	X
Derived							X	X	X

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Results from first stage of ignition estimation, logistic regression of PSPS events. Health risk and SES are both measured as indices, with 0 being least vulnerable and 100 being most vulnerable. A positive coefficient indicates that higher health risk (lower SES) circuits have a greater rate of ignitions. Robust standard errors are reported.

avoids the identification concern that when a firm conducts a PSPS, I do not observe whether an ignition would have occurred without that intervention. In Appendix C, I conduct a robustness exercise using the full set of data. I evaluate the coefficients assuming that each PSPS event would be an ignition, or that no PSPS event would be an ignition. The conclusions below still hold in both alternate specifications, suggesting that my findings hold regardless of any changes in the relationship between weather and ignition probability over time.

Table 5 shows the results from the first stage, logistic regression of ignitions on the vulnerability indices and additional controls. The pseudo-R squared value is relatively low, even for the model with both primary and derived weather covariates. Many of the coefficient estimates are statistically indistinguishable from 0. In PG&E circuits, higher health risk circuits are significantly (at $p = 0.001$ level) less likely to have an ignition and lower SES circuits are more likely to have an ignition. This finding is robust to including population and weather variables. At the $p = 0.01$ level, ignitions in SDG&E lines are positively correlated with higher vulnerability indices, although these relationships are not significant after controlling for weather factors.

Table 6 shows the results from the second stage, fitting a linear model to the log predicted probabilities from the first stage. These coefficients show the degree that vulnerability indices are associated with the log probability of ignition. I suppress the R^2 value from the second stage procedure, as this statistic does not incorporate the uncertainty from the first stage and may be misleading. Magnitudes and significance of second stage estimates are generally quite similar to the first stage results.

	(1)	(2)	(3)	(4)
SCE x Health	-0.00015	-0.00178	-0.00141	0.00151
	(0.00085)	(0.00264)	(0.00284)	(0.00291)
SCE x SES	-0.00299**	0.00244	0.00074	0.00028
	(0.00094)	(0.00262)	(0.00278)	(0.00279)
PG&E x Health	-0.01231***	-0.01110***	-0.01294***	-0.01290***
	(0.00045)	(0.00115)	(0.00125)	(0.00126)
PG&E x SES	0.01340***	0.01463***	0.01025***	0.00959***
	(0.00046)	(0.00110)	(0.00130)	(0.00133)
SDG&E x Health	0.03221***	0.01620*	0.01388	0.01264
	(0.00182)	(0.00723)	(0.00840)	(0.00857)
SDG&E x SES	0.01972***	0.01955**	0.01409*	0.01079
	(0.00181)	(0.00625)	(0.00624)	(0.00659)
Observations	14925285	7590153	7590153	7590153
Popuation		X	X	X
Primary			X	X
Derived				X

Table 6: Results from second stage of ignition probability regression. A positive coefficient indicates that higher health risk (lower SES) circuits have a greater rate of ignitions. Standard errors are computed using Appendix B. Stars indicate significance at the $p = 0.05(^{*}), 0.01(^{**}),$ and $0.001(^{***})$ levels.

Some patterns from the coefficient estimates in Table 6 are similar to those of the PSPS decisions, although less precisely estimated. Without controlling for weather variation, I find that lower SES circuits have higher rates of ignition in PG&E and SDG&E, and lower rates of ignition in SCE. I find that higher health risk circuits have higher rates of ignition in SDG&E, and lower rates in PG&E. Controlling for population and weather variation, the only significant associations that remain are that lower SES circuits in PG&E have higher rates of ignition and that higher health risk circuits in PG&E have lower rates of ignition. This is similar to the findings from Table 4, although there is greater uncertainty. This suggests that population and weather variation are able to explain much of the observed differences in ignitions between more and less vulnerable communities.

4.3 PSPS Costs

I use two proxies to find how vulnerability indices correlate with the utility's computed cost of PSPS. This cost is the value the utility uses when weighing the costs and benefits of a shutoff; it reflects the estimated size of the disruption from declaring a PSPS event. As discussed in Section 3.2, I use customer minutes interrupted (CMI) and number of customers impacted. Results using log of CMI

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Log CMI	Log CMI	Log CMI	Log CMI	Log Cust	Log Cust	Log Cust	Log Cust
SCE x Health	-0.00808*	-0.00780*	0.00871	0.00933	-0.00633*	-0.00606	0.0106*	0.0116*
	(0.00327)	(0.00384)	(0.00489)	(0.00506)	(0.00309)	(0.00364)	(0.00479)	(0.00491)
SCE x SES	0.0303***	0.0298***	0.0148**	0.0125*	0.0253***	0.0246***	0.0121*	0.0115*
	(0.00497)	(0.00502)	(0.00512)	(0.00521)	(0.00486)	(0.00493)	(0.00507)	(0.00521)
PG&E x Health	-0.0208***	-0.0207***	-0.0157***	-0.0164***	-0.0167***	-0.0170***	-0.0140***	-0.0135***
	(0.00275)	(0.00280)	(0.00286)	(0.00294)	(0.00255)	(0.00258)	(0.00261)	(0.00267)
PG&E x SES	0.0119**	0.0118**	0.00779*	0.0107**	0.00731*	0.00783*	0.00297	0.00372
	(0.00363)	(0.00369)	(0.00385)	(0.00384)	(0.00335)	(0.00340)	(0.00355)	(0.00355)
SDG&E x Health	-0.0108	-0.0315**	-0.0389*	-0.0303	-0.00472	-0.0274**	-0.0342*	-0.0297
	(0.00803)	(0.0104)	(0.0167)	(0.0185)	(0.00654)	(0.00856)	(0.0140)	(0.0164)
SDG&E x SES	0.00697	-0.00131	0.0132	0.0149	0.00375	-0.00582	0.00683	0.00754
	(0.00545)	(0.00620)	(0.0110)	(0.0126)	(0.00476)	(0.00531)	(0.00955)	(0.0117)
Observations	3270	3270	3270	3270	3270	3270	3270	3270
R^2	0.172	0.176	0.309	0.351	0.106	0.112	0.237	0.268
Population		X	X	X		X	X	X
Primary			X	X			X	X
Derived				X				X

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Results from linear regression of proxies to PSPS cost, CMI (columns 1-4) and number of customers (columns 5-8). A positive coefficient indicates that higher health risk (lower SES) circuits have higher average CMI or number of customers impacted. Robust standard errors are reported. Outcome variable values of 0 are omitted.

are shown in columns 1-4 of Table 7, and log of the number of customers impacted are in columns 5-8.

The patterns are generally similar between regressions, although both are relatively noisy. Estimates with log CMI are generally larger in magnitude and more precisely estimated than those using log number of customers. Without controlling for weather variation, there is a significant positive correlation between low SES and the cost proxy for SCE (p value < 0.001) and PG&E (p value 0.001 for CMI, 0.029 for customers). There is a significant negative correlation for the health risk index for PG&E (p value < 0.001). After controlling for weather variation, these observations are largely similar. The difference between estimates with no controls and all controls are not significant at the $p = 0.05$ except the "SCE x SES" row, where the association is no longer significant after including all controls.

Rules that determine the cost of PSPS may disadvantage low-SES or high-health risk populations if the number of customers is negatively correlated with these indices. This occurs regardless of whether the rules intend to discriminate based on these characteristics; that is, it is an example of statistical rather than taste-based discrimination (Guryan and Charles, 2013). The utility's decision rule places more weight on circuits with higher historical customer outages. If circuits with a higher share of vulnerable individuals are less impacted by shutoffs, the utility's rule calculates a lower cost from shutoffs in those circuits. These findings indicate that PG&E's decision rules may disadvantage high health risk circuits. Utilities can adjust their rules to avoid this potential discrimination, and some already are. SCE already scales part of their PSPS risk score by the size of populations with medical needs (Valdberg, Tozer, and Kilberg, 2021, p. 16), and SDG&E has plans to implement a similar program (SDG&E, 2021a, p. 30).

4.4 Implied Coefficient of Expected Damages

While I am not able to estimate the firm's expected damages from a wildfire, the results of the other estimation steps and the firm's decision-making rule allow me to infer how vulnerability indices influence the firm's expected damages from a wildfire. In Equation (2), I write a model of firms' decision making involving three parameters: σ_1 (the extent that vulnerability indices influence the firms' expected damages), σ_2 (the extent that vulnerability indices influence log probability of ignition), and σ_3 (the extent that vulnerability indices influence log cost of declaring PSPS). I am not able to estimate σ_1 directly, but estimate $\sigma_1 + \sigma_2 - \sigma_3$ in Section 4.1. With separate estimates of σ_2 and σ_3, I can infer a value of σ_1 as $\hat{\sigma}_1 = \{\sigma_1 + \sigma_2 - \sigma_3\} - \hat{\sigma}_2 + \hat{\sigma_3}$.

Table 8 shows the implied estimate of σ_1 given this model and my estimates of σ_2 and σ_3 from previous sections. I find standard errors by assuming that these problems are uncorrelated; I take the square root of the sum of squared standard deviations from other estimates. Overall, these estimates are quite noisy. I fail to reject that expected damages from fires are equitably estimated by utilities.

These implied values assume that utilities have no control over the duration of an outage or the number of customers impacted. This is supported by regulatory filings, which state that the size and duration of an outage are as small as permitted by weather and infrastructure conditions. This may not be accurate in practice. For example, I conclude that PG&E's expected damage is smaller for populations with higher health risk because there fewer customers are impacted in circuits with higher health risk. This observation would also be consistent with a model where firms face different costs for different populations, and set the length of an outage or the number of customers

	from CMI	from CMI	from CMI	from CMI	from Cust	from Cust	from Cust	from Cust
SCE x Health	-0.0113**	-0.00957	0.0109	0.00673	-0.00953**	-0.00784	0.0129*	0.00902
	(0.00377)	(0.00516)	(0.00607)	(0.0063)	(0.00361)	(0.00501)	(0.00599)	(0.00617)
SCE x SES	0.0277***	0.0217***	-0.0107	-0.0117	0.0226***	0.0166**	-0.0134*	-0.0128*
	(0.00544)	(0.00604)	(0.00641)	(0.00649)	(0.00535)	(0.00596)	(0.00637)	(0.00648)
PG&E x Health	-0.0278***	-0.0279***	-0.0137***	-0.0184***	-0.0237***	-0.0243***	-0.0121***	-0.0154***
	(0.00315)	(0.00336)	(0.00343)	(0.00359)	(0.00297)	(0.00318)	(0.00323)	(0.00338)
PG&E x SES	0.0106**	0.00824*	0.00679	0.0162***	0.00606	0.00428	0.00197	0.00927*
	(0.00388)	(0.00405)	(0.00436)	(0.00446)	(0.00362)	(0.00379)	(0.00409)	(0.00421)
SDG&E x Health	-0.0036	-0.00118	-0.0142	-0.0116	0.00252	0.00295	-0.00949	-0.0111
	(0.00927)	(0.0136)	(0.0198)	(0.0218)	(0.00801)	(0.0123)	(0.0176)	(0.0201)
SDG&E x SES	0.0119	0.0196*	0.0122	0.0226	0.00869	0.0151	0.00589	0.0153
	(0.00698)	(0.00961)	(0.0134)	(0.0151)	(0.00645)	(0.00906)	(0.0123)	(0.0143)
Population		X	X	X		X	X	X
Primary			X	X			X	X
Derived				X				X

Table 8: Inferred value of σ_1, or the coefficient of vulnerability indices on log of expected utility loss from a fire. Computed using coefficient estimates from previous regressions. Standard errors are the square root of sum of squares from previous regressions. Stars indicate significance at the $p = 0.05(^{*}), 0.01(^{**}),$ and $0.001(^{***})$ levels.

impacted to prioritize populations with higher health risk. I do not have the data necessary to falsify this model, although it would be of interest to distinguish between these potential models of firm conduct.

5 Conclusions

I find that PSPS is used more frequently in low-SES circuits among two of California's major utilities, and among higher health risk circuits in one of the major utilities. This finding is robust to controlling for weather variation. After controlling for weather variation, I find that ignitions are more frequent in low-SES circuits and in lower health risk circuits in PG&E, but otherwise do not find significant evidence. With my proxies for the firm's cost of declaring a PSPS event, I am unable to make strong claims about the degree that these vulnerability indices influence the utilities' estimated costs from declaring a PSPS shutoff or their expected damages from wildfires. I cannot reject that these costs are equitably distributed.

This work starts to explore a gap in the literature on empirically assessing the equity of adaptation mechanisms. More research is needed in this area more broadly, as well as to better understand the impacts of electric utilities' response to wildfire risk. This research agenda is challenging without better data about the firm's problem, particularly how the firm computes costs and benefits of

PSPS. These data would allow researchers to explore a broader range of research questions, such as the explorations of systematic bias in Obermeyer et al. (2019) or Rambachan (2021).

References

Abatzoglou, John (2013). "Development of gridded surface meteorological data for ecological applications and modelling". In: *International Journal of Climatology* 33.1, pp. 121–131.

Anderson, Sarah, Andrew Plantinga, Matthew Wibbenmeyer, et al. (2020). "Inequality in agency responsiveness: evidence from salient wildfire events". In: *Resources for the Future: Washington, DC, USA*, p. 36.

Athey, Susan, Julie Tibshirani, and Stefan Wager (2019). "Generalized random forests". In: *The Annals of Statistics* 47.2, pp. 1148–1178.

August, Laura, Komal Bangia, Laurel Plummer, Shankar Prasa, Kelsey Ranjbar, Andrew Slocombe, and Walker Wieland (2021). *CalEnviroScreen 4.0.* California Environmental Protection Agency, Office of Environmental Health Hazard Assessment.

Becker, Gary S (1957). *The economics of discrimination.* University of Chicago press.

Buolamwini, Joy and Timnit Gebru (2018). "Gender shades: Intersectional accuracy disparities in commercial gender classification". In: *Conference on fairness, accountability and transparency.* PMLR, pp. 77–91.

California Public Utilities Commission (2014). *Order instituting rulemaking to revise and clarify commission regulations relating to the safety of electric utility and communications infrastructure provider facilities. Decision 19-05-042.*

— (2019). *Decision adopting de-energization (Public Safety Power Shut-off) guidelines (Phase 1 Guidelines). Decision 19-05-042.*

Coggins, Shaugn, Lea Berrang-Ford, Keith Hyams, Poshendra Satyal, James Ford, Jouni Paavola, Ingrid Arotoma-Rojas, and Sherilee Harper (2021). "Empirical assessment of equity and justice in climate adaptation literature: A systematic map". In: *Environmental Research Letters* 16.7, p. 073003.

D'Evelyn, Savannah M et al. (2022). "Wildfire, Smoke Exposure, Human Health, and Environmental Justice Need to be Integrated into Forest Restoration and Management". In: *Current environmental health reports*, pp. 1–20.

Gianfrancesco, Milena A, Suzanne Tamang, Jinoos Yazdany, and Gabriela Schmajuk (2018). "Potential biases in machine learning algorithms using electronic health record data". In: *JAMA internal medicine* 178.11, pp. 1544–1547.

Guliasi, Leslie (2021). "Toward a political economy of public safety power shutoff: Politics, ideology, and the limits of regulatory choice in California". In: *Energy Research & Social Science* 71, p. 101842.

Guryan, Jonathan and Kerwin Kofi Charles (2013). "Taste-based or statistical discrimination: The economics of discrimination returns to its roots". In: *The Economic Journal* 123.572, F417–F432.

Hill, L L, R Blythe, E M Krieger, A Smith, A McPhail, and S B C Shonkoff (2020). "The public health dimensions of California wildfire and Wildfire Prevention, Mitigation, and Suppression". In: *Technical Report; Physicians and Scientists for Healthy Energy.*

Holmes, Thomas P, Robert J Huggett, and Anthony L Westerling (2008). "Statistical analysis of large wildfires". In: *The economics of forest disturbances.* Springer, pp. 59–77.

Ichimura, Hidehiko and Whitney K Newey (2022). "The influence function of semiparametric estimators". In: *Quantitative Economics* 13.1, pp. 29–61.

IPCC (2022). *Climate Change 2022: Impacts, Adaptation, and Vulnerability. Working Group II Contribution to the IPCC Sixth Assessment Report.* Ed. by H.-O. Pörtner et al. Cambridge University Press.

Jain, Piyush, Sean CP Coogan, Sriram Ganapathi Subramanian, Mark Crowley, Steve Taylor, and Mike D Flannigan (2020). “A review of machine learning applications in wildfire science and management”. In: *Environmental Reviews* 28.4, pp. 478–505.

Jann, Ben (2020). “Influence functions continued. A framework for estimating standard errors in reweighting, matching, and regression adjustment”. In.

Khan, Shakeeb and Elie Tamer (2009). “Inference on endogenously censored regression models using conditional moment inequalities”. In: *Journal of Econometrics* 152.2, pp. 104–119.

Lang, Kevin and Ariella Kahn-Lang Spitzer (2020). “Race discrimination: An economic perspective”. In: *Journal of Economic Perspectives* 34.2, pp. 68–89.

McFadden, D. (1973). “Conditional Logit Analysis of Qualitative Choice Behaviour”. In: *Frontiers in Econometrics*. Ed. by P. Zarembka. New York, NY, USA: Academic Press New York, pp. 105–142.

Mehrabi, Ninareh, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan (2021). “A survey on bias and fairness in machine learning”. In: *ACM Computing Surveys (CSUR)* 54.6, pp. 1–35.

Newey, Whitney K and Daniel McFadden (1994). “Large sample estimation and hypothesis testing”. In: *Handbook of econometrics* 4, pp. 2111–2245.

Niemi, Ernest G. and Kristin S Lee (2001). “Wildfire and poverty: an overview of the interactions among wildfires, fire-related programs, and poverty in the Western States”. In: The Center for Watershed and Community Health, Mark O. Hatfield School of Government, Portland State University.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan (2019). “Dissecting racial bias in an algorithm used to manage the health of populations”. In: *Science* 366.6464, pp. 447–453.

Ojerio, Ryan S (2008). “Equity in Wildfire Risk Management: Does Socioeconomic Status Predict Involvement in Federal Programs to Mitigate Wildfire Risk?” MA thesis. Department of Planning, Public Policy & Management, University of Oregon.

Oster, Emily (2019). “Unobservable selection and coefficient stability: Theory and evidence”. In: *Journal of Business & Economic Statistics* 37.2, pp. 187–204.

Pacific Gas and Electric Company (2020). *2020 risk assessment and mitigation phase report*. Tech. rep.

— (June 2021). *2021 Wildfire mitigation plan – revised*. Tech. rep.

Parikh, Ravi B, Stephanie Teeple, and Amol S Navathe (2019). “Addressing bias in artificial intelligence in health care”. In: *Jama* 322.24, pp. 2377–2378.

Plantinga, Andrew J, Randall Walsh, and Matthew Wibbenmeyer (2022). “Priorities and effectiveness in wildfire management: evidence from fire spread in the western United States”. In: *Journal of the Association of Environmental and Resource Economists* 9.4, pp. 603–639.

Rambachan, Ashesh (2021). “Identifying prediction mistakes in observational data”. In: *Working Paper, Harvard University Department of Economics.*

Rhodes, Noah, Lewis Ntaimo, and Line Roald (2020). “Balancing wildfire risk and power outages through optimized power shut-offs”. In: *IEEE Transactions on Power Systems* 36.4, pp. 3118–3128.

San Diego Gas and Electric Company (Feb. 2021a). *2020-2022 Wildfire mitigation plan update*. Tech. rep.

— (2021b). *San Diego Gas and Electric Company: risk assessment mitigation phase*. Tech. rep.

Satyal, Poshendra, Morten Fibieger Byskov, and Keith Hyams (2021). “Addressing multi-dimensional injustice in indigenous adaptation: the case of Uganda’s Batwa community”. In: *Climate and Development* 13.6, pp. 529–542.

Sheller, Mimi and Yolanda M Leon (2016). "Uneven socio-ecologies of Hispaniola: Asymmetric capabilities for climate adaptation in Haiti and the Dominican Republic". In: *Geoforum* 73, pp. 32–46.

Short, K C (2014). *A spatial database of wildfires in the United States, 1992-2011.*

Short, Karen C (2021). *Spatial wildfire occurrence data for the United States, 1992-2018.*

Smit, Barry and Olga Pilifosova (2003). "Adaptation to climate change in the context of sustainable development and equity". In: *Sustainable Development* 8.9, p. 9.

Southern California Edison Company (June 2021). *2021 Wildfire mitigation plan update (revision).* Tech. rep.

— (2022). *Southern California Edison Company: risk assessment mitigation phase.* Tech. rep.

Taylor, Steve W, Douglas G Woolford, CB Dean, and David L Martell (2013). "Wildfire prediction to inform fire management: statistical science challenges". In: *Statistical Science* 28.4, pp. 586–615.

Thomas, David SG and Chasca Twyman (2005). "Equity and justice in climate change adaptation amongst natural-resource-dependent societies". In: *Global environmental change* 15.2, pp. 115–124.

Valdberg, Anna, Andrea Tozer, and Elena Kilberg (2021). *Southern California Edison Company's (U 338-E) public safety power shutoff post-event report for November 21, 2021 de-energization event.* Tech. rep.

Wibbenmeyer, Matthew and Molly Robertson (2022). "The distributional incidence of wildfire hazard in the western United States". In: *Environmental Research Letters* 17.6, p. 064031.

Wigtil, Gabriel, Roger B Hammer, Jeffrey D Kline, Miranda H Mockrin, Susan I Stewart, Daniel Roper, and Volker C Radeloff (2016). "Places where wildfire potential and social vulnerability coincide in the coterminous United States". In: *International Journal of Wildland Fire* 25.8, pp. 896–908.

Wong-Parodi, Gabrielle (2020). "When climate change adaptation becomes a "looming threat" to society: Exploring views and responses to California wildfires and public safety power shutoffs". In: *Energy Research and Social Science* 70.

Xi, DD, Stephen W Taylor, Douglas G Woolford, and CB Dean (2019). "Statistical models of key components of wildfire risk". In: *Annual Review of Statistics and Its Application* 6.1, pp. 197–222.

A Predicting Fire Size

I use public data to attempt to predict fire size given weather covariates. From 1992-2018, comprehensive records of fire size are available from the US Forest Service (K C Short, 2014; Karen C Short, 2021). From 2019-2021, I include records from the National Interagency Fire Service.[4] Records include the date, fire size, and latitude and longitude of ignition. The final database includes 240,239 records within California. I then merge these data with my weather observations from GridMET.

I model the problem of predicting catastrophic fires both as a regression and classification problem. To predict fire size, I regress the log of fire size against the full set of weather variables from GridMET, as well as yearly fixed effects and fixed effect terms per utility's service area. For classification trials, I use three definitions of "large fire": top 0.02 quantile (my definition), larger than 300 acres[5], and larger than 500 acres.[6]. In each classification trial, I weight each observation by the inverse frequency of its class to predict the relatively rare event of a large fire. I consider a linear set of weather variables, linear regression with interactions between weather variables, and random forests with 5-folds cross validation.

Table 9 summarizes the results of these trials. For classification trials, I report the specificity (share of negative outcomes that are correctly predicted) and sensitivity (share of positive outcomes that are correctly predicted) of each prediction method, for each "large fire" definition. For regression, I report the R^2 value. Figure 5 shows the scatter plots of predicted fire size vs. actual fire size.

Overall, these results indicate poor performance at predicting fire size. I have limited ability to extrapolate fire size, meaning I cannot construct informative bounds on the missing data as required to identify counterfactuals in Rambachan (2021).

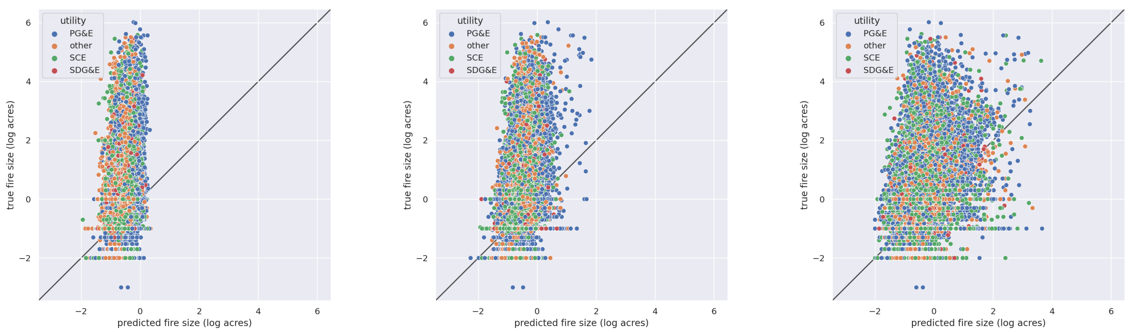
[4] From `https://data-nifc.opendata.arcgis.com/datasets/nifc::wfigs-current-wildland-fire-perimeters/about`, accessed 11 January 2022.

[5] Definition from `https://www.nps.gov/olym/learn/management/upload/fire-wildfire-definitions-2.pdf`, accessed 1 April 2022.

[6] Definition from Holmes, Huggett, and Westerling (2008).

	Regression R squared	Fire size $\geq$ 300 acres Specificity	Sensitivity	Fire size $\geq$ 500 acres Specificity	Sensitivity	Fire size $\geq$ top 2% Specificity	Sensitivity
Linear	0.06311	0.6907	0.6455	0.7017	0.6624	0.6691	0.6115
Linear Interacted	0.08927	0.723	0.6612	0.7302	0.6893	0.6888	0.6271
Random Forest	0.08446	0.02007	0.9994	0.01766	0.9994	0.02622	0.9991

Table 9: Results from random forest and linear regression at predicting large fires.



(a) Linear (b) Linear interacted (c) Random forest

Figure 5: Predicted vs actual fire size, using various regression methods.

B Standard Errors for Ignition Probability

In this section, I derive the asymptotic variance of the two-step estimation procedure from Section 3.1. I write the problem as a two-step M estimation, and apply the method from Newey and Daniel McFadden (1994) to find asymptotic variance after accounting for error in estimating g.

I rewrite Equation (7) as a least squares problem. Let $m(\theta, X_i, Z_i, g) := -(\log \hat{\pi}(X_i, Z_i; g) - \sigma_2 X_i - \tilde{\pi}(Z_i))^2$, emphasizing the dependence of predicted probabilities $\hat{\pi}$ on the first-step function g. The set of parameters θ includes σ_2 and the parameters defining the function $\tilde{\pi}$. Then:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} m(\theta, X_i, Z_i, g) \tag{8}$$

I can then express the asymptotic variance of estimating θ, as long as $\hat{g}$ is asymptotically linear. The form comes from applying the Mean Value Theorem to the differences $\sqrt{N}(\hat{\theta} - \theta_0)$ and $\sqrt{N}(\hat{g} - g_0)$.

$$\begin{aligned} &\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V), \text{ where} \\ &V = H^{-1}\mathbb{E}\left[\left\{\frac{\partial m(\theta, X_i, Z_i, g)}{\partial \theta} + M\psi(X_i, Z_i)\right\}\left\{\frac{\partial m(\theta, X_i, Z_i, g)}{\partial \theta} + M\psi(X_i, Z_i)\right\}'\right] H^{-1} \end{aligned} \tag{9}$$

where H is the Hessian matrix of m: $H = \mathbb{E}\left[\frac{\partial^2 m(\theta, X_i, Z_i, g)}{\partial\theta\partial\theta'}\right]$, M is the matrix of cross derivatives of m: $M = \mathbb{E}\left[\frac{\partial^2 m(\theta, X_i, Z_i, g)}{\partial\theta\partial g'}\right]$, and $\psi(X_i, Z_i)$ is the influence function of g for an observation $\{X_i, Z_i\}$.

With linear approximations to the functions θ and g, it is straightforward to derive these expressions. Let $W_i = \{X_i, Z_i\}$ be the column vector of all covariates for unit i. Then let $\beta = \theta$ so that $\sigma_2 X_i + \tilde{\pi}(Z_i) = W_i'\beta$, and let ρ be a parameter vector so that $g(X_i, Z_i) = W_i'\rho$. Recall that with the logistic assumption, $\hat{\pi}(W_i; \rho) = (1 + \exp(-W_i'\rho))^{-1}$. I write $\hat{\pi}_i := \hat{\pi}(W_i; \rho)$ Then:

$$m(\beta, W_i, \rho) = -(\log\hat{\pi}_i - W_i'\beta)'(\log\hat{\pi}_i - W_i'\beta) \tag{10}$$

$$\frac{\partial m(\beta, W_i, \rho)}{\partial\beta} = 2W_i(\log\hat{\pi}_i - W_i'\beta) \tag{11}$$

$$H = -2\mathbb{E}[W_i W_i'] \tag{12}$$

$$M = 2\mathbb{E}\left[W_i \frac{\exp(-W_i'\rho)}{1 + \exp(-W_i'\rho)} W_i'\right] = 2\mathbb{E}[W_i(1 - \hat{\pi}_i)W_i'] \tag{13}$$

Following Jann (2020), I derive the influence function for logistic regression after writing the problem as a Generalized Method of Moments estimator:

$$\psi(W_i) = \left(\mathbb{E}[W_i\hat{\pi}_i(1 - \hat{\pi}_i)W_i']\right)^{-1} W_i(y_i - \hat{\pi}_i) \tag{14}$$

Plugging these into Equation (9), I can determine the asymptotic variance of my estimator. Note that this framework allows for more general semiparametric estimation, as long as $\hat{g}$ is asymptotically linear. This allows estimators like random forests or general double machine learning techniques; see Athey, Tibshirani, and Wager (2019) for advice on deriving the influence function of a random forest and Ichimura and Newey (2022) for influence functions of general semiparametric functions.

C Ignitions regression with alternate sample

In the main text, I estimate ignition probability using data from years where firms do not declare PSPS, to avoid a potential data censoring problem. As discussed in Section 3.1, this choice if the relationship between weather variables and ignition probability is not consistent between years when utilities do and do not use PSPS. As a robustness exercise, I include results using the full time period, with two different assumptions on the observed ignitions and shutoffs: that, absent a shutoff, each circuit with a PSPS event would either have an ignition (Table 10 and Table 11) or no ignition (Table 12 and Table 13). Overall, these results suggest that my conclusions in the main analysis are robust to including ignition data from the full sample.

This provides suggestive evidence about the partially identified set that contains the true parameter. To fully characterize that set, I could enumerate all possible potential realizations of the missing data and repeat the estimation procedure for each potential outcome. Due to the immense computational cost of such a procedure, I only repeat the two-stage estimation procedure for these two scenarios.

Table 10 and Table 11 show the results from the first- and second-stage estimation (respectively), using the full sample and treating each missing value as a true positive. The findings that are statistically significant (at the $p = 0.001$ level) from Table 6 in the main analysis are also significant in these regressions: low-SES circuits in PG&E have higher rates of ignition, as do lower health risk circuits. This regression finds additional significant associations, although many of these are not robust to an alternate assumption on the missing data.

Table 12 and Table 13 show the results from the first- and second-stage estimation (respectively), using the full sample and treating each missing value as a true negative. Again, the statistically significant findings from the main analysis are confirmed in these regressions. Many of the additional significant associations from treating each missing value as a true positive are not significant in this exercise, although both find that lines in low-SES circuits in SDG&E are significantly more likely to have an ignition.

	(1) ignition	(2) ignition	(3) ignition	(4) ignition	(5) ignition	(6) ignition	(7) ignition	(8) ignition	(9) ignition
Health Risk Index	-0.000153	-0.0123^{***}	0.0322^{***}	0.00128	-0.0117^{***}	0.0243^{***}	0.00238	-0.0115^{***}	0.0189^{***}
	(0.00113)	(0.000615)	(0.00272)	(0.00141)	(0.000617)	(0.00303)	(0.00151)	(0.000691)	(0.00374)
SES Index	-0.00299^{*}	0.0134^{***}	0.0197^{***}	-0.00359^{**}	0.0129^{***}	0.0253^{***}	-0.0148^{***}	0.00918^{***}	0.0100^{***}
	(0.00134)	(0.000583)	(0.00292)	(0.00138)	(0.000575)	(0.00255)	(0.00163)	(0.000746)	(0.00290)
Observations	3120696	9879282	1925307	3120696	9879282	1925307	3120696	9879282	1895814
Pseudo R^2	0.034	0.029	0.045	0.035	0.030	0.069	0.198	0.168	0.334
Utility	SCE	PG&E	SDG&E	SCE	PG&E	SDG&E	SCE	PG&E	SDG&E
Population				X	X	X	X	X	X
Primary							X	X	X
Derived							X	X	X

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: First stage ignition results, using the full sample, with PSPS results counted as ignitions. This is the assumption that all censored results would have been true positives.

	(1)	(2)	(3)	(4)
SCE x Health	-0.00015	0.00128	-0.00204	-0.00224
	(0.00113)	(0.00141)	(0.00131)	(0.00143)
SCE x SES	-0.00299*	-0.00359**	-0.01008***	-0.01056***
	(0.00134)	(0.00138)	(0.00161)	(0.00162)
PG&E x Health	-0.01231***	-0.01170***	-0.01136***	-0.01241***
	(0.00062)	(0.00062)	(0.00067)	(0.00069)
PG&E x SES	0.01340***	0.01285***	0.00807***	0.00965***
	(0.00058)	(0.00057)	(0.00071)	(0.00074)
SDG&E x Health	0.03221***	0.02432***	0.02337***	0.02043***
	(0.00272)	(0.00303)	(0.00372)	(0.00374)
SDG&E x SES	0.01972***	0.02532***	0.01335***	0.01204***
	(0.00292)	(0.00255)	(0.00277)	(0.00290)
Observations	14925285	14925285	14850777	14889258
Popuation		X	X	X
Primary			X	X
Derived				X

Table 11: Results from second stage of ignition probability regression, using the full sample, with PSPS results counted as ignitions. Standard errors are adjusted to account for error in estimating the probabilities; derivation is given in Appendix B. Stars indicate significance at the $p = 0.05(^{*}), 0.01(^{**}),$ and $0.001(^{***})$ levels.

	(1) ignition	(2) ignition	(3) ignition	(4) ignition	(5) ignition	(6) ignition	(7) ignition	(8) ignition	(9) ignition
Health Risk Index	0.00144	-0.0119***	0.0179***	0.00460*	-0.0117***	0.0133**	0.00849***	-0.0134***	0.00875
	(0.00155)	(0.000769)	(0.00430)	(0.00187)	(0.000776)	(0.00461)	(0.00207)	(0.000866)	(0.00547)
SES Index	0.00298	0.0167***	0.0205***	0.00178	0.0166***	0.0224***	-0.00347	0.0123***	0.0132**
	(0.00180)	(0.000759)	(0.00404)	(0.00182)	(0.000761)	(0.00376)	(0.00199)	(0.000925)	(0.00406)
Observations	3120696	9879282	1867158	3120696	9879282	1867158	3120696	9879282	1845942
Pseudo R^2	0.007	0.007	0.013	0.008	0.007	0.019	0.052	0.061	0.088
Utility	SCE	PG&E	SDG&E	SCE	PG&E	SDG&E	SCE	PG&E	SDG&E
Population				X	X	X	X	X	X
Primary							X	X	X
Derived							X	X	X

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

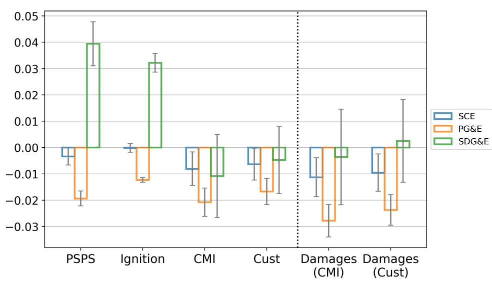
Table 12: First stage ignition results, using the full sample, with no PSPS results counted as ignitions. This is the assumption that all censored results would have been true negatives.

	(1)	(2)	(3)	(4)
SCE x Health	0.00144	0.00460*	0.00177	0.00400*
	(0.00155)	(0.00187)	(0.00198)	(0.00202)
SCE x SES	0.00298	0.00178	0.00073	0.00051
	(0.00180)	(0.00182)	(0.00196)	(0.00197)
PG&E x Health	-0.01185***	-0.01175***	-0.01385***	-0.01391***
	(0.00077)	(0.00078)	(0.00086)	(0.00086)
PG&E x SES	0.01669***	0.01660***	0.01268***	0.01256***
	(0.00076)	(0.00076)	(0.00090)	(0.00092)
SDG&E x Health	0.01787***	0.01328**	0.01025	0.01004
	(0.00430)	(0.00461)	(0.00542)	(0.00546)
SDG&E x SES	0.02049***	0.02238***	0.01569***	0.01445***
	(0.00404)	(0.00376)	(0.00375)	(0.00404)
Observations	14867136	14867136	14844334	14844604
Popuation		X	X	X
Primary			X	X
Derived				X

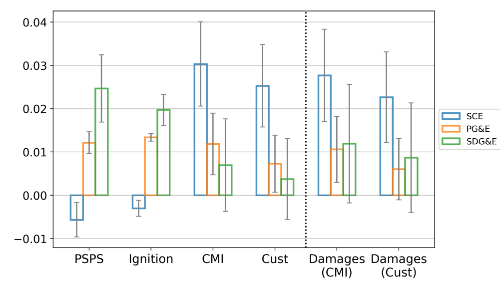
Table 13: Results from second stage of ignition probability regression, using the full sample, with no PSPS results counted as ignitions. Standard errors are adjusted to account for error in estimating the probabilities; derivation is given in Appendix B. Stars indicate significance at the $p = 0.05(^{*}), 0.01(^{**}),$ and $0.001(^{***})$ levels.

D Additional Coefficient Plots

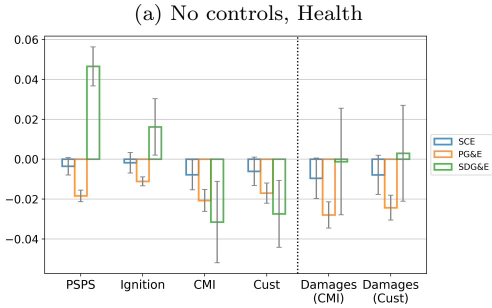
In the main text, I include a visualization of coefficient estimates using all controls (weather and population). In this section, I include additional visualizations for the other specifications.



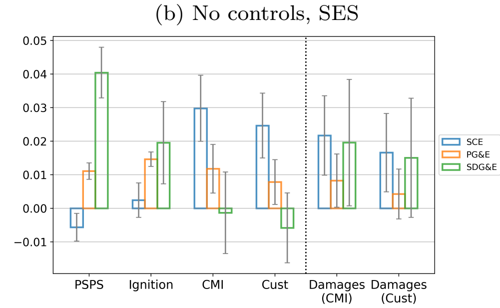
(a) No controls, Health



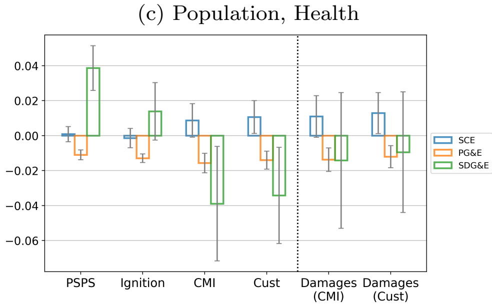
(b) No controls, SES



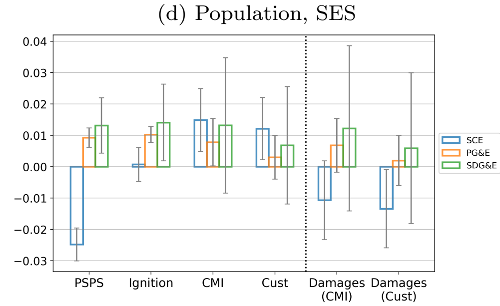
(c) Population, Health



(d) Population, SES



(e) Primary weather, Health



(f) Primary weather, SES

Figure 6: Collected coefficient estimates of health risk or SES index on various outcomes, using varying sets of controls. Each group of plots on the x axis show coefficient estimates for a given outcome variable. Damages (to the right of dotted line) are implied from estimates to the left of dotted line and my model. Error bars show ± 1.96 times the standard error. Each group of plots is ordered SCE, PG&E, SDG&E.