Paying for Electricity in California: How Residential Rate Design Impacts Equity and Electrification

Online Appendix

Severin Borenstein, Meredith Fowlie and James Sallee*

April 3, 2023

Abstract

.

*Borenstein: Haas School of Business and Energy Institute at Haas, University of California, Berkeley, CA 94720-1900, severinborenstein@berkeley.edu. Fowlie: Department of Agricultural & Resource Economics and Energy Institute at Haas, University of California, Berkeley, CA 94720, fowlie@berkeley.edu. Sallee: Department of Agricultural & Resource Economics and Energy Institute at Haas, University of California, Berkeley, CA 94720, sallee@berkeley.edu. For outstanding research assistance and very helpful comments, we thank Marshall Blundell. For valuable comments and discussions, we thank Andy Campbell. This research was funded in part by Next 10 (www.next10.org). Online appendix is available at http://??

This document provides technical supporting information for the report titled *Paying for Electricity in California: How Residential Rate Design Impacts Equity and Electrification.* In the first section, ...:https://github.com/marshallblundell/PfE.

1 Appendix 1: Solar imputation

The billing data we use for our analysis includes the net consumption of households with rooftop solar systems. To understand how billing would change under rate reforms that change the compensation structure for solar generation, we need an estimate of the gross consumption (net consumption plus rooftop generation).

The billing data do not include information on the size of a household's solar system. We estimate system size by matching each solar customer to an installed system from the Tracking the Sun database provided by Lawrence Berkeley National Lab (LBNL) that has the size, location and installation date for all residential systems in the state.[1]

We match systems in the LBNL data to households in the billing data based on zip code, timing of installation, and the relationship between system size and household consumption. More specifically, both the LBNL data and the billing data have zip code. The billing data lists an interconnection date, and the LBNL data has a month of installation. Due to potential differences in how these dates get recorded, it is possible that the a system has a different date in the two data sources. As a result, we attempt to match based on quarter, but we allow some flexibility in how closely dates are aligned, as described below.

We explored several methods for matching. A primary challenge is that there are some very large systems. Assigning them to households based only on zip code and timing of installation sometimes matches implausibly large systems to what appear to be households with a small load. As a result, we added an element of matching based on household electricity load.

In the billing data, to asses a household's load, we construct a maximum observed monthly net consumption as follows. For each billing cycle, we calculate the observed net consumption per day. For each household, we find the maximum of this over all the billing cycles.

For households who installed solar before 2016 (44.5% of systems), we only observe net consumption with the system installed. For these households, within each zip code, we rank order solar PV system sizes with largest maximum observed net consumption and assign them to systems from the LBNL data in rank order. That is, we assign the largest pre-2016 system within a zip code to the largest net consumer in that zip code who installed solar before 2016, and so on.

For households that installed solar in 2016 or after, we observe some billing cycles with consumption before they installed solar. For these households, we determine whether a household is

[1] More details on Tracking the Sun are available here: https://emp.lbl.gov/tracking-the-sun.

a plausible match for a given system by calculating its average consumption before installation. A household is "eligible" to match with a system so long as its average consumption before solar installation is at least 18% of the system's capacity. This is based on an approximation of the typical capacity factor in California and a rule of thumb that installers have not sized systems to exceed consumption.

For households that installed solar in 2016 or after, we identify each system in the LBNL data and look for a household in the same zip code to match it with. Within a zip code, we start with the largest system in the LBNL data, and look for an eligible household in the same zip code in the billing data. We first look for an eligible household whose interconnection date in the billing data is in the same quarter as the LBNL installation date. The household with the largest pre-solar average consumption is assigned if there is an eligible match. Otherwise, we extend the search to the quarter before and after, and assign a match as above. If there is still no eligible match, we look for a match in the months two quarters before or after the LBNL installation date. If no household is eligible, then we leave the LBNL system unassigned. We then proceed to the next largest system in the zip code within the LBNL data, and so on. A household that is matched to an eligible system already is not available to match with a later system. Household systems in the billing data that are not matched to any system in the LBNL data at the end of this process are assigned the median capacity of systems in their zip code in the year of their installation.

Once we assign a system size to each household, we estimate generation for that system using a calculator provided by PV Watts to estimate a capacity factor, which we then multiply by system size. For tilt and azimuth we use the zip code average tilt and azimuth of systems provided in the LBNL data.[2] Our estimate of gross consumption for a household is the observed net consumption plus generation.

2 Appendix 2: Inferring household-level income

We are interested in analyzing the incidence of the current retail electricity rate regime in California. To do this kind of analysis credibly, we need reliable estimates of household-level income. We do not observe income in our electricity billing data. To impute income measures, we estimate the empirical relationship between variables we observe in our utility data (i.e., census block group, CARE participation, and household electricity consumption) and income measures we can obtain from other sources.

When individual data are unavailable, so-called *ecological regression* techniques are sometimes

[2]More details on PV Watts are available at https://pvwatts.nrel.gov/. For systems not matched to a system in the LBNL data we use the median capacity of systems in that zip code and the average tilt and azimuth of systems in that zip code. For zip codes for which all systems in the LBNL data do not have tilt and azimuth, we use the latitude of the zip code centriod for the tilt and 180 degrees for the azimuth of systems in that ZIP Code.

used to estimate individual relationships between variables using group-level data. In our context, this amounts to regressing mean or median income—which is readily observed at the zip code or census block group level—on the corresponding group-average electricity consumption and CARE participation rates. This is common practice in the most relevant literature, but concerns with this approach have been documented, including in Borenstein (2012).

An important limitation of the ecological regression approach is that the statistical relationship obtained at the group level will not hold at the individual level if there is within-group variation in electricity consumption that is confounded with factors that determine income. In our case, as we will show, there is significant within-group variability in household electricity consumption and in income. Moreover, in survey data, household income and electricity consumption are positively correlated within groups. Thus, we cannot assume that the regression coefficients we estimate using aggregate data identify the underlying household-level relationships.

To avoid relying on potentially misleading ecological regressions for our income imputation exercise, we leverage two additional data sources. The first is the 2019 American Community Survey (ACS). The household income measures in the ACS include pretax cash income of all people 15 years and older in the household. The ACS has an annual sample size of about 3.5 million addresses. The 5-year ACS estimates are constructed using data collected over a 5-year time period. Using the 5-year data, we can construct statistically reliable distributions of household income at the CBG level.[3]

If we assumed a perfect rank order correlation between electricity consumption and household-level income within each CBG, we could simply sort households within a CBG by gross electricity consumption and assign each a corresponding position in the CBG-specific income distribution. However, we know from surveys that collect information about household income and electricity consumption that this rank order correlation is far from perfect (see, for example, the Residential Energy Consumption Survey). Thus, invoking this assumption would lead us to overstate the progressivity of the current retail electricity rate structure. On the other hand, assuming zero correlation between gross electricity consumption and household-level income within each CBG would understate the progressivity of the current rate structure, because surveys indicate there is some positive correlation.

To combine the income distribution information from the ACS with the consumption distribution from the billing data for every CBG, we need some way of matching households across the two data sets. We do this by estimating the observed relationship between income and consumption in data from California's Residential Appliance Saturation Survey (RASS). This is a stratified random survey of California households which provides a rare opportunity to directly observe

[3]These distributions are based on approximately a 15 percent Census population survey sample over 5 years.

the empirical relationship between household-level income and household-level electricity consumption. Electricity consumption data in the RASS are verified through a customer match with utilities, which makes consumption measures much more reliable than some other data sources like the Consumer Expenditure Survey. In addition, this is a larger sample within California than the most prominent federal survey on home energy consumption, the Residential Energy Consumption Survey.

In a first step, we use the RASS survey data to empirically estimate the relationships between income, electricity consumption, and CARE program participation. The complication, as we explain below, is that the sample from the RASS is too small to estimate this relationship within CBG based on the RASS alone. We overcome this limitation by combining information from the RASS, the billing data and the ACS. We then use this imputed income, together with estimated CARE participation rates by income septile and the CBG-specific income distributions, to assign each household to one of seven income categories. In what follows, we briefly summarize the RASS survey data and explain the income imputation process in detail.[4]

2.1 California's 2019 RASS Survey

The 2019 RASS survey collects household-level information about appliance ownership, electricity consumption, rooftop solar adoption, and a categorical measure of self-reported household income. RASS respondents are asked to "check the range that best describes your household's total annual income." Table 1 summarizes the responses to this question using data from households located in the service territories of the three investor-owned utilities.

The overall response rate for the 2019 RASS survey was 9 percent. Of these respondents, Table 1 shows that 72 percent answered the income question. The subset of surveyed households that report income is not perfectly representative of all California households. However, as Figures 1a: 3a show, the distribution of household incomes reported in the RASS sample (blue bars) are quite similar to the income distributions we observe in the ACS data (red bars).

We merge the RASS data with utility-provided data on household-level electricity consumption, CARE program participation, and census block group location. Table 1 shows that most, but not all, households that report income also report annual electricity consumption. For the subset

[4]Borenstein (2012) dealt with a similar challenge of assigning incomes to electricity customers recognizing that income heterogeneity would likely be correlated with consumption heterogeneity within CBGs. Whereas that paper calibrated a simple correlation coefficient to the overall means by income category in the RASS, in this paper we use the full richness of the RASS data to estimate empirically the association of income with consumption and CARE participation accounting more precisely for the income categorization in the ACS.

Table 1: RASS Responses by Income Category

Category	Income range	Households	Share of Total	Share Reporting Electricity Consumption
1	\$0-\$10k	721	2.1	82.5
2	\$10k-\$19k	1,649	4.7	86.2
3	\$20-\$25k	1,384	4.0	86.0
4	\$25k - \$50k	4,182	12.0	85.8
5	\$50k - \$75k	3,973	11.4	80.6
6	\$75k - \$100k	3,204	9.2	77.3
7	\$100-\$150k	4,240	12.2	76.1
8	\$150k-\$175k	1,558	4.5	76.6
9	\$175-\$200	1,005	2.9	76.2
10	\$200-\$250	1,279	3.7	72.9
11	$> \$250$	1,829	5.3	68.8
12	Prefer not to answer	5,024	14.4	.
97	Missing	4,754	13.7	.
Total		34,802	100	.

Notes: This table summarizes responses of RASS households located in the service territories of the three investor-owned utilities.

of households responding to both questions, Table 2 summarizes CARE participation rates and annual electricity consumption (net of solar PV generation) by income category.

As expected, CARE participation rates are much higher among lower-income groups. However, not all eligible (based on income) CARE households participate. At the other end of the income distribution, some presumably ineligible households in higher income categories are CARE customers. This suggests that CARE participation is a coarse proxy for low income. The table also shows that average annual net electricity consumption from the grid does increase with income (although not monotonically). The table shows significant variation in net electricity consumption within income categories. We also find significant variation in both income and electricity consumption within census block groups.

2.2 Household-level income estimation

We use RASS survey data collected from residential customers of three investor-owned utilities to estimate the empirical relationship between household-level income and electricity consumption. We specify the following estimating equation which allows this relationship to differ systematically across CARE and non-CARE customers:

Table 2: CARE participation and net grid electricity consumption by income category

Category	Income range	CARE participation rate.	Annual net grid electricity consumption (kWh)	Number of households
1	\$0-\$10k	88%	4,360 (2,733)	595
2	\$10k-\$19k	87%	4,782 (3,096)	1,422
3	\$20-\$25k	81%	5,046 (3,451)	1,190
4	\$25k - \$50k	56%	5,245 (2,971)	3,588
5	\$50k - \$75k	27%	5,578 (3,434)	3,201
6	\$75k - \$100k	15%	5,846 (3,479)	2,478
7	\$100-\$150k	9%	5,867 (3,221)	3,198
8	\$150k-\$175k	7%	6,200 (3,310)	1,193
9	\$175-\$200	7%	6,605 (3,927)	766
10	\$200-\$250	5%	6,341 (3,330)	932
11	$> \$250$	3%	7,306 (4,874)	1,259
Total		33%	5,678 (3,459)	19,822

Notes: This table summarizes responses of RASS households located in the service territories of the three investor-owned utilities.

$$Y^*_{ig} = \gamma_g + \beta C_{ig} + \alpha_1 CARE_{ig} + \alpha_2 CARE_{ig} \cdot C_{ig} + \epsilon_{ig}, \tag{1}$$

where Y^*_{ig} denotes the annual income earned by household i in location g, where location can be either a 5-digit zip code, census tract, or census block group (CBG). The γ_g are location fixed effects which captures the average household income within a location. The C_{ig} denotes annual gross electricity consumption at the household level for household i who resides in location g. For households with rooftop solar, we add simulated solar PV generation to net grid electricity consumption (see Appendix 1). The $CARE_{ig}$ indicator switches on if household i participates in the CARE program. The last term, ϵ_{ig}, is an error term. Equation (1) uses within-location variation in income, electricity consumption, and CARE participation to identify the α and β parameters.

Given those parameters, we can use the observed consumption, geographic location, and CARE participation status of households in our billing data to estimate their income.

We face two main challenges in estimating the α and β parameters. One challenge is that we do not directly observe household income in the RASS data. Instead, households indicate the income range they fall within. In one specification, we use the midpoint of the reported income range to proxy for Y^*_{ig}. However, this approach will generate biased estimates because variation in the household income is not fully captured by this categorical measure. Our preferred specification is an ordered probit model with known income thresholds. This approach more accurately captures the relationship between the latent income values we do not observe, Y^*, and the income category Y we do observe. The probit model assumes that the error term ϵ_{ig} has a normal distribution with mean zero and variance, σ^2, where σ is a parameter to be estimated.

To illustrate the intuition behind this probit model, consider a simple example in which there are only three income categories. Let κ_j define the j interval threshold points:

$$\begin{aligned} Y &= 1 \quad \text{if } Y^* \leq \kappa_1 \\ Y &= 2 \quad \text{if } \kappa_1 \leq Y^* \leq \kappa_2 \\ Y &= 3 \quad \text{if } \kappa_2 \leq Y^*. \end{aligned}$$

We estimate the probability that the unobserved variable Y^* falls within each threshold. We define an indicator I_i such that:

$$I_i = j \text{ if } \kappa_{j-1} < I^*_i \leq \kappa_j.$$

The probability that household i falls into the income category j that we observe can be defined as:

$$Prob(I = j|x) = \left[\Phi\left(\frac{\kappa_j - \beta' X}{\sigma}\right) - \Phi\left(\frac{\kappa_{j-1} - \beta' X}{\sigma}\right)\right], \tag{2}$$

where $\beta' X = \beta C_{ig} + \alpha_1 CARE_{ig} + \alpha_2 CARE_{ig} \cdot C_{ig} + \gamma_g$. [5]

We search for the α, β, and σ parameters that maximize the likelihood function: $L(\alpha, \beta, \sigma) = \prod_{i=1}^{N} \prod_j P_{ij}^{y_{ij}}$, where y_{ij} is an indicator denoting the income category j observed for household i. Because we know the income threshold values, both the β and the residual variance parameter σ can be identified.

The other main challenge is that we wish to focus on CBG as our location variable, so as to leverage the maximum available information about income from Census data. The RASS sample,

[5] Note that the γ_g can be defined in terms of CBG group means. We calibrate the CBG mean income using census data and the CBG mean electricity consumption and average CARE participation using billing data. This allows us to estimate the latent income equation without having to estimate CBG fixed effects directly.

however, is too small to estimate CBG fixed effects directly–many CBG's would have only one or two households. Instead, we use the ACS and our billing data to calculate the mean income, mean consumption and mean CARE participation rate of households within each CBG. Then, rather than estimating equation 1 directly, we demean all variables by CBG and run regressions on the demeaned data. The demeaned regression and the original equation yield numerically identical coefficients when the demeaning is done within sample. The difference in our case is that we calculate means for each variable using an additional data source.

Table 3-5 summarizes results from estimating ecological regressions using different levels of aggregation, estimating the latent income equation using OLS (defining income using the mid-point values), and estimating the probit model with known income thresholds. Where CBG fixed effects are indicated, the regression is run on the demeaned data as described above. There is one table for each utility.

Table 3: Estimated conditional correlations between income, electricity consumption, and CARE participation (PG&E customers)

Variable	Ecological (zip) (1)	Ecological (census tract) (2)	Ecological (block group) (3)	RASS (OLS) (4)	RASS (Probit model) (5)
Annual electricity consumption (C)	5.02	13.38	13.71	5.81	5.23
	(2.25)	(1.15)	(0.93)	(0.45)	(0.28)
CARE participation	-125,562	-20,756	-438	-54,050	-54,741
	(43,165)	(17,950)	(12,741)	(3,453)	(3,599)
CARE x C	-15.00	-11.07	-15.03	-2.65	-2.08
	(6.63)	(2.43)	(2.05)	(0.52)	(0.51)
σ	-	-	-	-	71,548
					(649)
FE	–	ZIP	ZIP	CBG	CBG
R^2	0.63	0.49	0.42	0.13	0.04
N	780	4,740	10,751	7,259	7,259

This table reports estimates for the latent income equation parameters. All regressions that use billing data are population weighted. Regressions that use RASS data are unweighted. Standard errors are clustered at the level of the group fixed-effects included in the estimating equation. The R-squared reported for the probit estimation is a pseudo-R-squared.

Table 4: Estimated conditional correlations between income, electricity consumption, and CARE participation (SCE customers)

Variable	Ecological (zip) (1)	Ecological (census tract) (2)	Ecological (block group) (3)	RASS (OLS) (4)	RASS (Probit model) (5)
Annual electricity consumption (C)	10.48	10.32	11.56	4.58	4.20
	(1.53)	(0.79)	(0.95)	(0.49)	(0.24)
CARE participation	1,155	-46,053	-37,120	-41,075	-44,498
	(24,001)	(13,186)	(13,830)	(3,166)	(3,271)
CARE x C	-27.02	-8.59	-8.58	-1.72	-1.15
	(4.02)	(1.82)	(1.75)	(0.52)	(0.48)
σ	-	-	-	-	57,390
					(6,545)
FE	–	ZIP	ZIP	CBG	CBG
R^2	0.71	0.54	0.46	0.12	0.02
N	534	4,834	10,179	6,545	6,545

This table reports estimates for the latent income equation parameters. All regressions that use billing data are population weighted. Regressions that use RASS data are unweighted. Standard errors are clustered at the level of the group fixed effects included in the estimating equation. The R-squared reported for the probit estimation is a pseudo-R-squared.

The estimates reported in Column (1) capture the relationships between averages taken over each zip code. At this level of aggregation, our regression coefficients are likely picking up the effects of confounding factors (e.g., climate, housing stock) that vary across zip codes. In the regressions reported in columns (2) and (3), data are aggregated to the level of census tract and census block group level, respectively. In these regressions, we can include zip code fixed effects to control for factors that vary across zip codes (on average). With these ecological regressions, we should still be concerned about aggregation bias. We know from the ACS data and utility billing data that income, electricity consumption, and CARE participation varies significantly within zip codes. Columns (1) through (3) use only census and billing data.

Column (4) reports results from the OLS regression estimated using household-level RASS data. Across all three utilities, we find that the electricity consumption coefficient identified using within-census block group variation is smaller than the electricity consumption coefficients we estimate using group-level data.

Table 5: Estimated conditional correlations between income, electricity consumption, and CARE participation (SDG&E customers)

Variable	Ecological (zip) (1)	Ecological (census tract) (2)	Ecological (block group) (3)	RASS (OLS) (4)	RASS (Probit model) (5)
Annual electricity consumption (C)	11.23	11.04	12.01	7.04	6.66
	(2.81)	(1.79)	(1.51)	(0.91)	(0.59)
CARE participation	-27,534	-45,366	-36,035	-17,467	-14,044
	(44,867)	(16,396)	(14,353)	(6,337)	(4,955)
CARE x C	-27.42	-6.14	-7.58	-2.75	-3.22
	(9.15)	(2.78)	(2.05)	(1.27)	(0.85)
σ	-	-	-	-	62,470
					(980)
FE	–	ZIP	ZIP	CBG	CBG
R^2	0.77	0.54	0.48	0.09	0.02
N	114	1,003	2,362	2,352	2,352

This table reports estimates for the latent income equation parameters. All regressions that use billing data are population weighted. Regressions that use RASS data are unweighted. Standard errors are clustered at the level of the group fixed effects included in the estimating equation. The R-squared reported for the probit estimation is a pseudo-R-squared.

Our preferred estimates are the probit model estimates reported in column (5). These estimates explicitly account for the categorical nature of household-level reported income. Estimates range from 4.2 (SCE) to 6.7 (SDGE). Recall that this equation is estimated using residualized household-level data. These coefficient estimates imply that an increase (reduction) of 1 kWh above (below) the CBG average annual electricity consumption is associated with an increase (decrease) of \$4.2-\$6.6 dollars in annual income. Across the three IOUs, we find that CARE participation (conditional on electricity consumption) is associated with a significant reduction below the CBG-average income. We also find that CARE participation attenuates the relationship between electricity consumption and income.

Having estimated the parameters of Equation 1, we impute an income value for each household in the billing data. Our estimate of σ, which is the standard deviation of the distribution of unexplained variation in latent income, plays an important role in this imputation exercise. For each household in the data, we take a random draw from this distribution to calibrate household-specific values of ϵ_{ig}. The probit assumes that ϵ_{ig} is normally distributed. Thus, an estimated σ

of \$71,548 (the estimate for PG&E) implies that the error we draw for a given household will be between -140,234 (=1.96 times \$71,548) and 140,234 95% of the time, with a heavier concentration at values closer to zero. Intuitively, within CBG, consumption and CARE participation explain only a modest fraction of income variation, so much variation is left to the error term.

2.3 Assigning households to income septiles

Having imputed income values for each household in the billing data, we could simply use these values as our best estimate of household income. Given the magnitude of our estimates of σ, our estimates of income are statistically imprecise. And this direct imputation approach would not incorporate the rich information we have from the ACS which characterizes the distribution of household income within each CBG using a large sample.

To leverage all of the information we have available to us, we develop an algorithm that assigns each household a place in the corresponding CBG income distribution in a way that both minimizes the distance between the CBG-level income rankings (implied by our imputed income values) and matches the utility-income-septile specific CARE participation rates we estimate using the RASS data. This algorithm is designed to assign each household in the billing data to one of the seven income septiles.[6] The final assignment closely matches the income ranking implied by our imputed income values (within a CBG) and the income septile-specific CARE participation rates we observe in RASS data.

The algorithm consists of several steps. We introduce additional notation to describe this process in detail. For ease of exposition, we do not include utility subscripts. All of the following are calibrated separately for each of the three IOUs.

- q_g: The count of customers in the billing data living in CBG g.
- q_{CARE_g}: The count of CARE customers in the billing data living in CBG g.
- q_{jg}: The count of customers in each income septile according to the ACS.
- w_{mm} : a weight that represents the number of customers per account ($w_{mm} > 1$ for master-metered customers).
- w_g: a weight for a customer assigned to multiple CBGs. A large majority of customers assigned to one CBG: $w_g = 1$.
- q_{jg}: The number of customers in CBG g that fall within income septile j.

[6] There are eleven income categories represented in the RASS data. There are 16 income categories in the ACS data. We use the cut-off points that are common to both RASS ad Census to define 7 income categories.

- α_j: The CARE participation rate in income septile j calibrated using RASS data.

- r : The ratio of total CARE participation in the billing data and total CARE participation observed in the RASS data.

- α_j^*: Adjusted CARE participation rate (to match billing data). Thus, α_j^* = $\alpha_j r$.

- θ_g: A group-specific inflation factor that scales each α_j^* up or down to match the total number of CARE households we observe in the billing data. Thus, $\theta_g = q_{CARE,g} / \sum_j (q_{j,g}\alpha_j^*)$.

- q_{CAREgj}. The counts of CARE customers for each income septile implied by our initial assignment. Thus, $q_{CAREgj} = \theta_g \alpha_j^* q_{j,g}$.

The initial income assignment implied by our imputed income values puts more CARE customers assigned in the lowest income category than there are customers in that income category. So we need to re-assign CARE customers to better align our imputed and observed participation rates.

Starting with the first income septile, we search for the scale factor, β_1^* which reduces the number of CARE customers in the lowest income septile such that the number of customers allocated to this septile 1, $q{*}_{CARE,g,1}$, matches the overall participation rate in septile 1: $\sum_g (q_{CAREg1}^*) / \sum_g (q_{g1}) = \alpha_1^*$. Note that, within any single CBG, we do not want to allocate more CARE customers than the number of CARE customers we actually observein that CBG. Thus, we require that $q{*}_{CAREgj} < q_{gj}$. We also want to avoid under-allocating CARE customers such that we leave more CARE customers to allocate than slots remain in other septiles. We thus require that $q_{CARE,g} - \sum_1^j (q_{CAREgj}^*) \leq q_g - \sum_1^j (q_{gj})$. Imposing these two requirements, we search over β_1^* values to find the value that minimize the absolute difference between α_1^* and $\sum_g q_{CARE,g,1}^* / \sum_g q_{g,1}$.

Next, we adjust numbers of CARE customers in remaining septiles to match the total number of CARE customers we observe in a CBG. More precisely, we compute an adjustment factor $\theta_g = (q_{CAREg} - q_{CAREg1}^*) / \sum_{j>1} (q_{CAREgj})$. We then update q_{CAREgj} for $j > 1$, the counts of CARE customers for each income septile implied by our initial assigment, by multiplying by θ_g.

We repeat the above for septiles $j = 1 - 6$. Once customers have been allocated to the sixth septile, the allocation for the seventh is also complete as it holds the remaining customers. Having re-allocated CARE customers across all income septiles, we check to make sure that hat we have not allocated more CARE customers to a septile than that CBG's income distribution allows, and that the total number of CARE customers in a CBG matches what we actually observe in the utility billing data. Having confirmed that these constraints are met, we make a final adjustment to reweight master-meter accounts (reflecting the fact that there are multiple households behind master meters).

Figures 1a: 3a compare the results of this income imputation exercise against the income distributions we observe in RASS and the Census data, respectively. In general, the income distributions are quite similar across the Census data (red), the RASS data (blue), and our imputed income values (green). Our income imputation exercise is disciplined by both the income distributions we observe at the CBG level and the CARE participation rates we observe in the RASS data. Thus, the income distributions generated by our assignment algorithm do not match the Census distribution exactly.

Figures 1b: 3b compare CARE participation by income septile across the RASS sample and the billing data (wherein income has been imputed). Our algorithm matches the PG&E and SCE CARE program participation shares quite closely.

The CARE participation rate in the SDGE RASS sample is 53.7%, almost twice the population rate of 28.1% that we observe in the billing data. This indicates that the RASS survey over-sampled CARE customers. Unmodified, our algorithm matches α_j^*, an adjusted CARE participation rate (to match billing data), where $\alpha_j^* = \alpha_j r$. For SDG&E, this step would result in a participation rate in the lowest income septile of below 50%, which is implausible given the CARE eligibility criteria and the participation rates we observe in the PG&E and SCE RASS samples. We modify our algorithm by setting $\alpha_j^* = \alpha_j$, and matching the unadjusted participation rates in the RASS sample. The unadjusted participation rate is in line with adjusted rates in the PG&E and SCE samples.

3 Appendix 3: Calculations with alternative values of the social cost of carbon

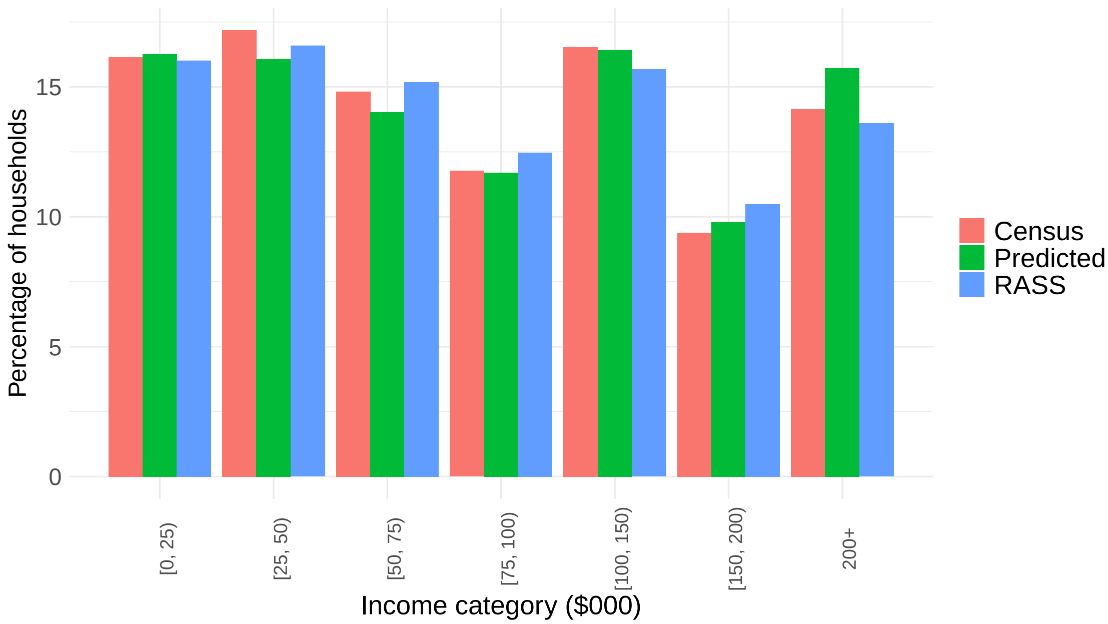
Our report uses $50 per ton for the externality costs of CO2 emissions, commonly known as the social cost of carbon. This is consistent with federal government guidelines as of 2021, but there is a great deal of uncertainty around that number.

A higher social cost of carbon will shrink the effective electricity tax because a higher cost of pollution implies a higher social marginal cost. Changing the social cost of carbon changes nearly all of the numbers we calculate. To give a sense of how different values might influence our analysis, this appendix provides estimates of all of our main figures that are sensitive to the social cost of carbon using both $50 per ton (our base case, for reference) and $100 per ton.

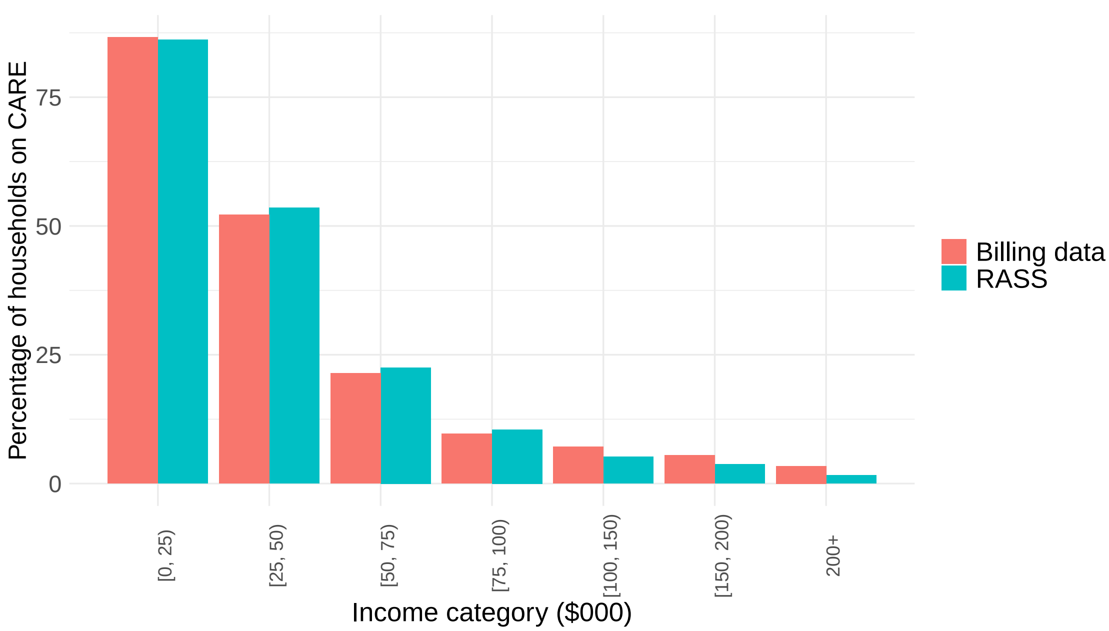
While all of our numbers change, the main point of the paper and key conclusions are generally robust to much higher pollution costs. Because we estimated the marginal emissions rate to be around 0.4 tonnes per MWh, a $50 increase in the social cost of carbon adds about 2 cents per kWh to SMC. Every $100 per tonne increase in the social cost of carbon shrinks the electricity

Figure 1: PG&E Household Income Distributions and CARE Participation

(a) Household income distribution (2019)



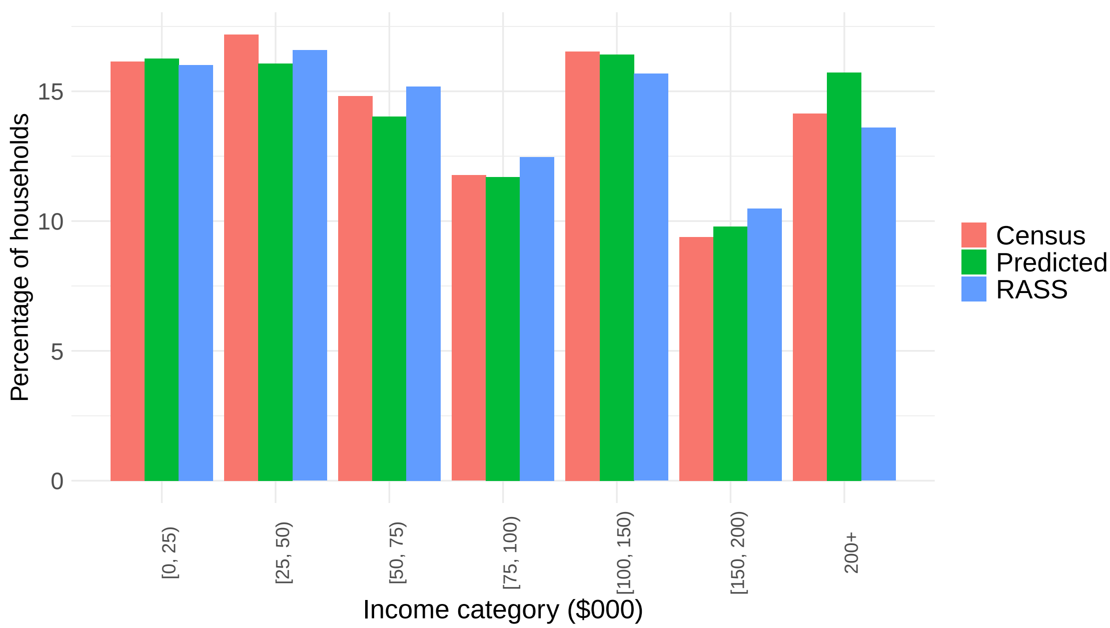
(b) CARE participation by income septile



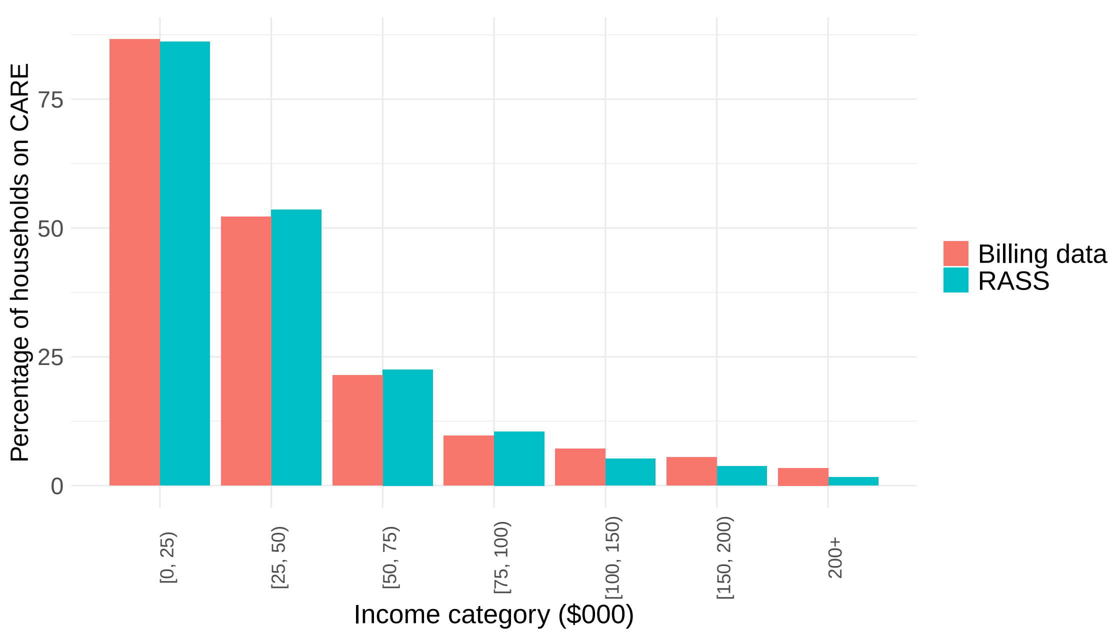
The top panel shows how households in PG&E service territory are distributed across income septiles using the Census data, our predicted income values, and the RASS data. The bottom panel shows CARE program participation rates among PG&E households by income septile using our predicted income values (in red) and using RASS survey responses (in blue).

Figure 2: SCE Household Income Distributions and CARE Participation

(a) SCE Household income distribution (2019)



(b) SCE CARE participation by income septile

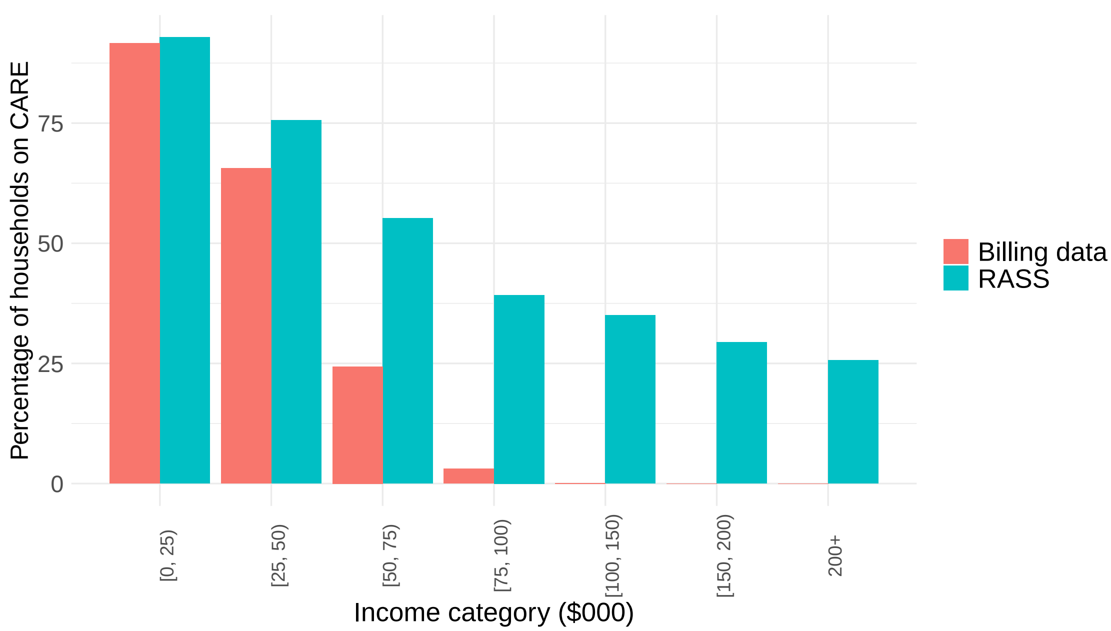


The top panel shows how households in SCE service territory are distributed across income septiles using the Census data, our predicted income values, and the RASS data. The bottom panel shows CARE program participation rates among SCE households by income septile using our predicted income values (in red) and using RASS survey responses (in blue).

Figure 3: SDG&E Household Income Distributions and CARE Participation

(a) SDG&E Household income distribution (2019)

(b) SDG&E CARE participation by income septile



The top panel shows how households in SDG&E service territory are distributed across income septiles using the Census data, our predicted income values, and the RASS data. The bottom panel shows CARE program participation rates among SDG&E households by income septile using our predicted income values (in red) and using RASS survey responses (in blue).

tax by about 4 cents per kWh, so electricity remains overpriced even at very high estimates of the social cost of carbon given that the effective tax we find in 2019 is 12 to 20 cents per kWh across California IOUs (using a $50/tonne SCC).

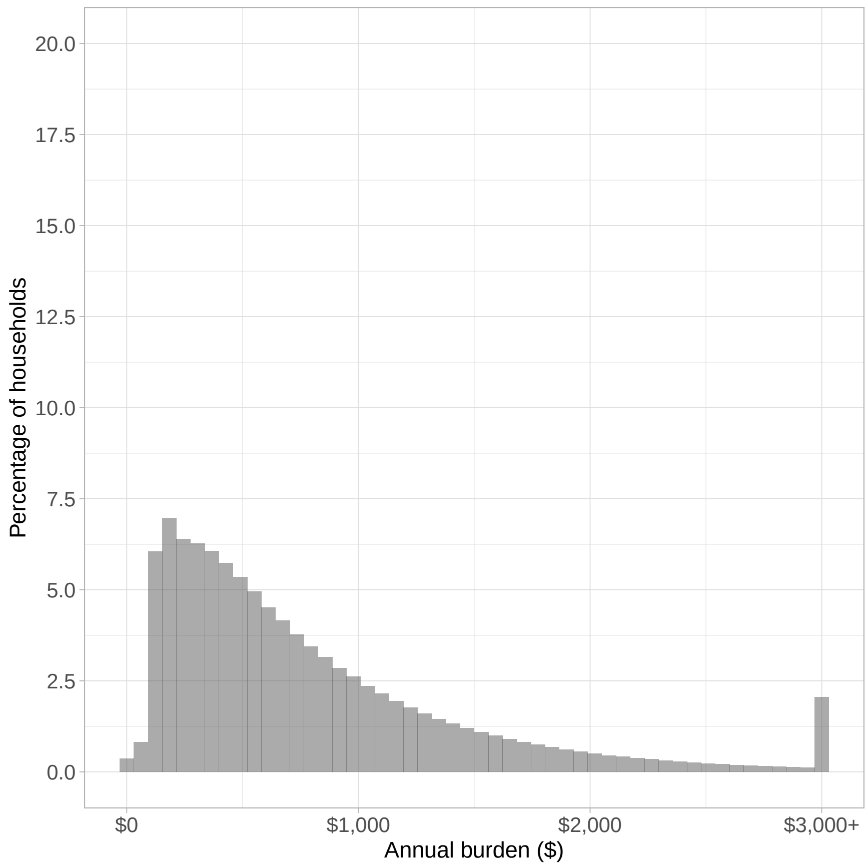
Importantly, rising electricity prices that are not associated with rising SMC have the opposite effect—they increase the effective electricity tax. Thus, if consensus moves towards using a higher social cost of carbon, we believe it is likely that this will be more than offset in the near term by rising electricity prices due to wildfire and other components of residual costs, such that the effective electricity tax in the near future is likely to rise, not fall. In fact, rate changes since 2019, the year on which we focus in the analysis, suggest the electricity tax has increased in just the last few years. Thus, for instance, although figure 5 suggests that under a $100 social cost of carbon, the electricity tax would be more progressive than the gas tax and about as progressive as a sales tax, that would only be true if there were no increase in the residual cost burden after 2019.

References

Borenstein, Severin. 2012. "The Redistributional Impact of Non-Linear Electricity Pricing." American Economic Journal: Economic Policy, 4(3): 56–90.

Figure 1: PG&E Annual Residual Cost Burden (2019)

(a) $50 Social Cost of Carbon



(b) $100 Social Cost of Carbon

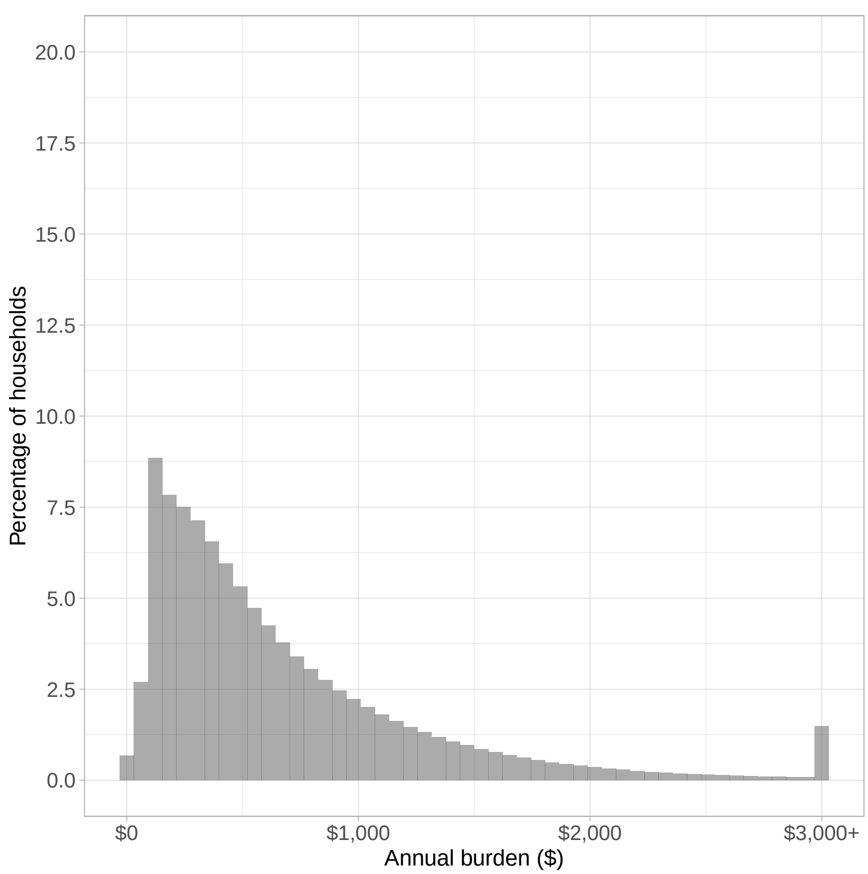
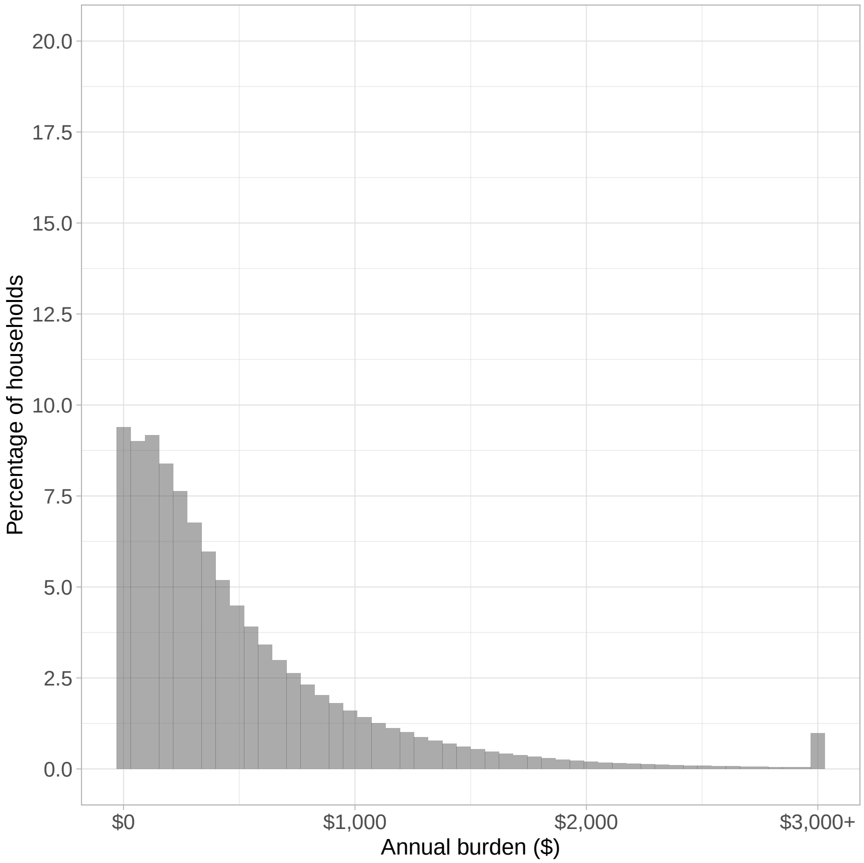


Figure 1: SCE Annual Residual Cost Burden (2019)

(a) $50 Social Cost of Carbon



(b) $100 Social Cost of Carbon

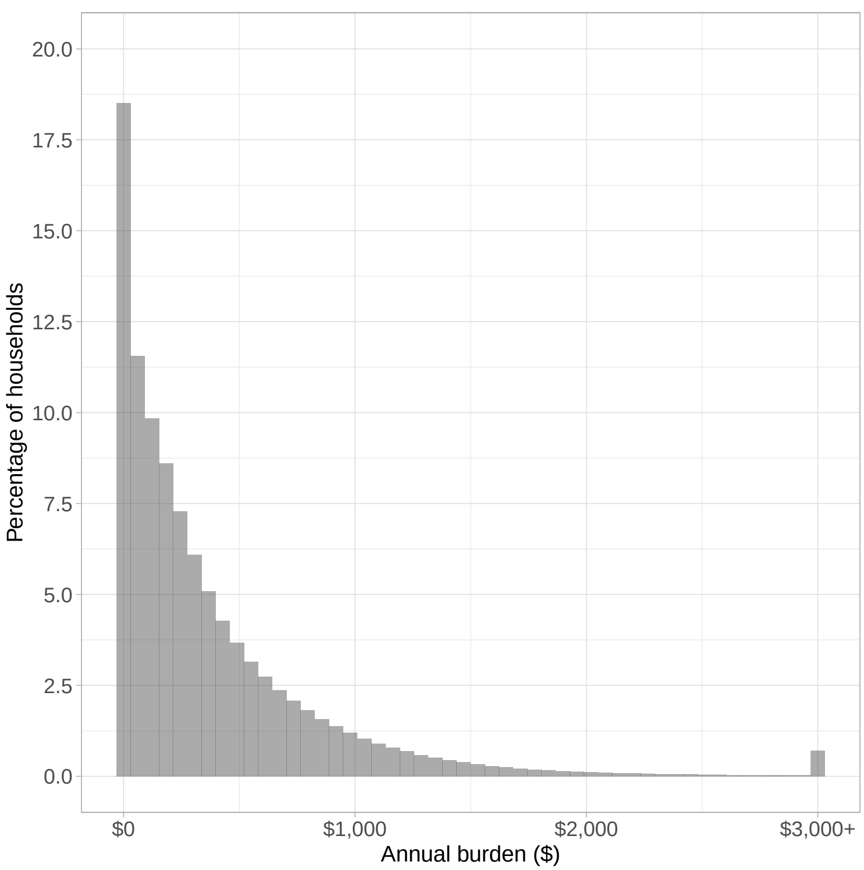
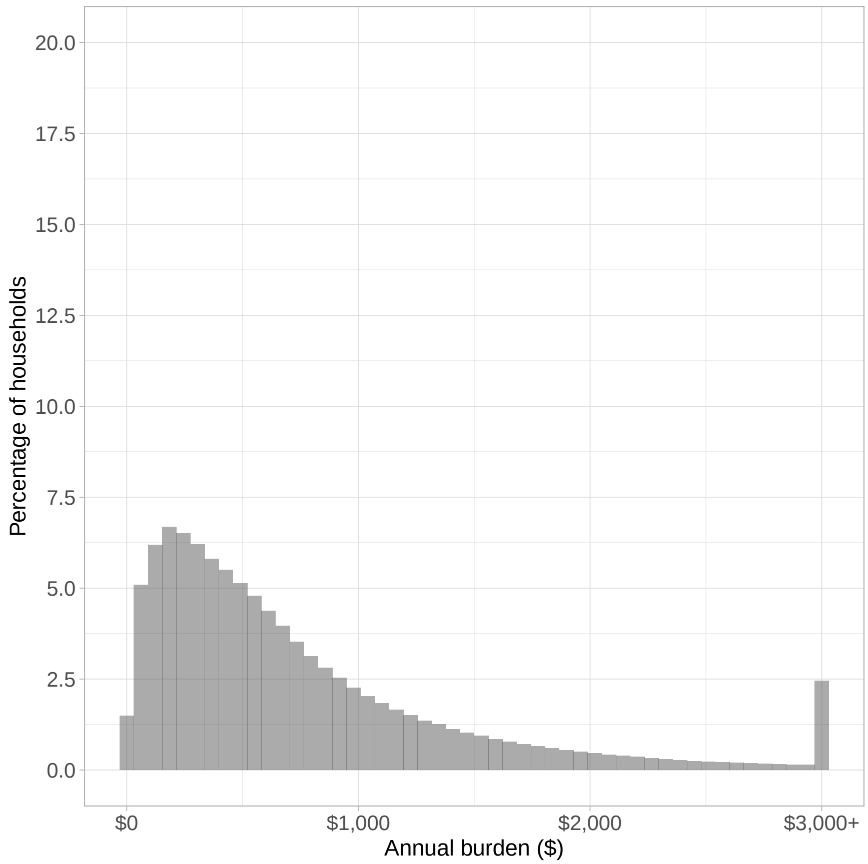


Figure 1: SDG&E Annual Residual Cost Burden (2019)

(a) $50 Social Cost of Carbon



(b) $100 Social Cost of Carbon

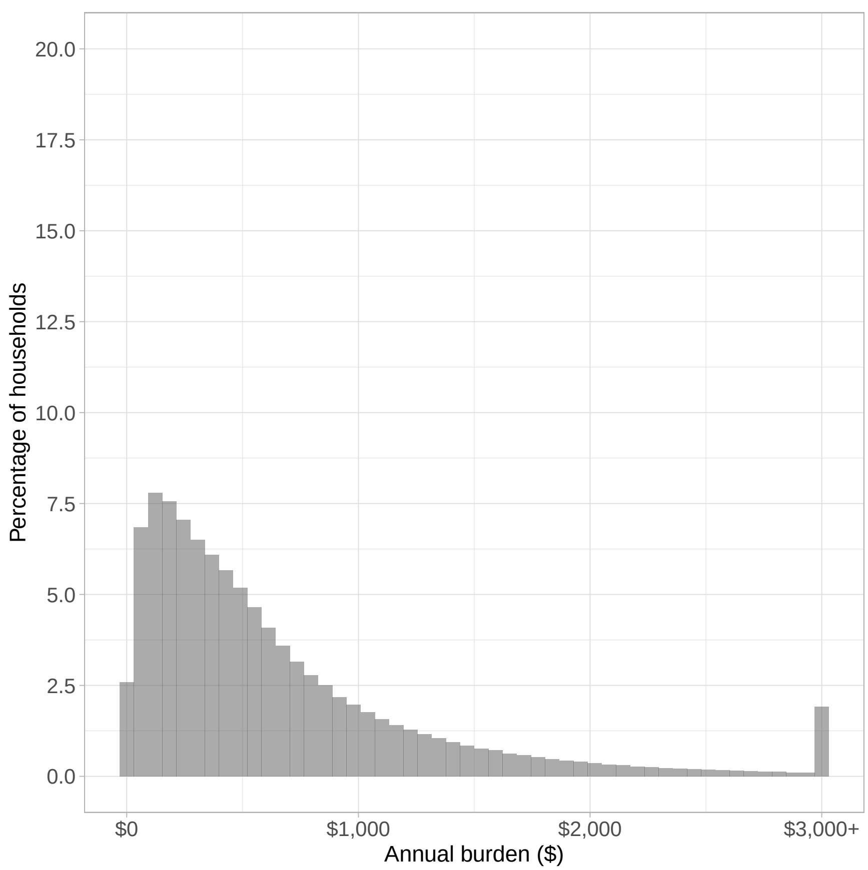
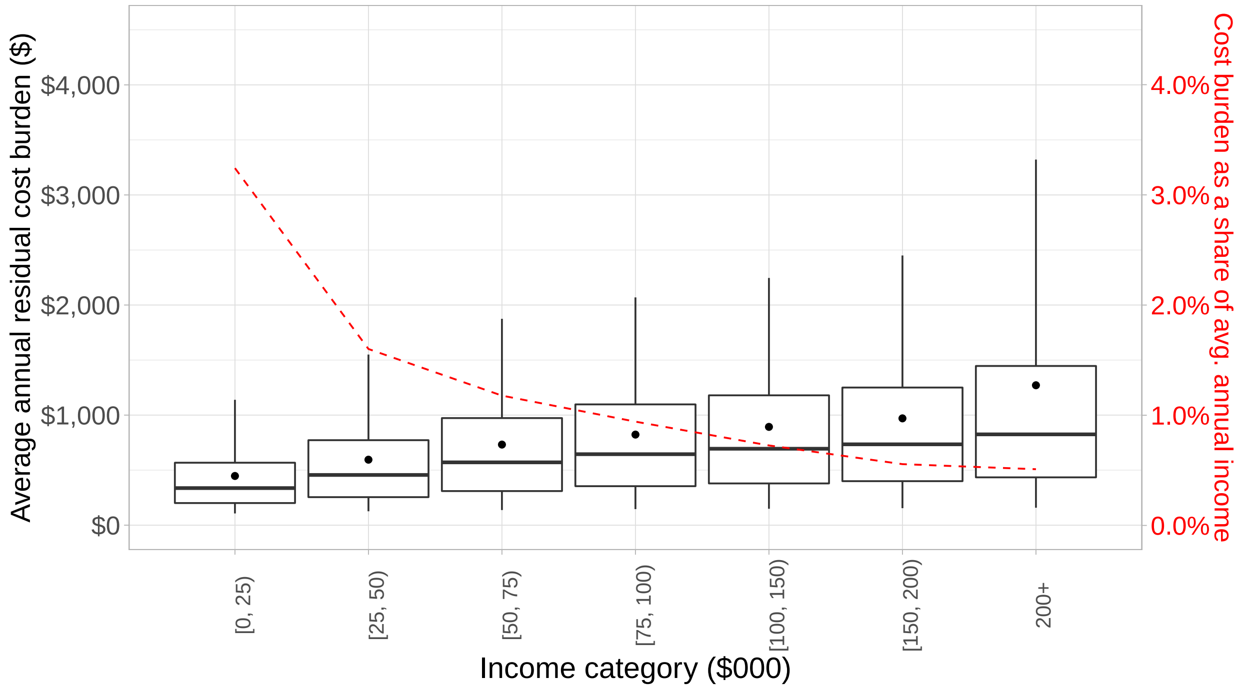


Figure 3: PG&E Annual Residual Cost Burden (2019)
Annual bill increase from charging current rates rather than SMC

(a) $50 Social Cost of Carbon



(b) $100 Social Cost of Carbon

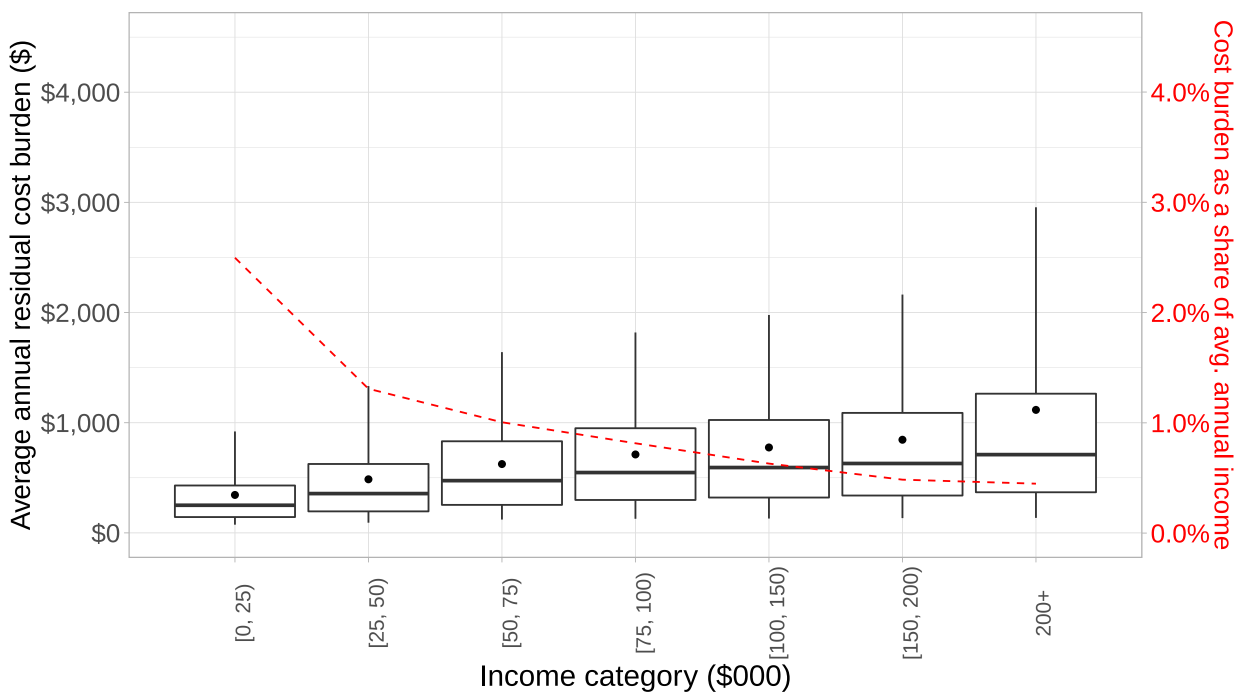
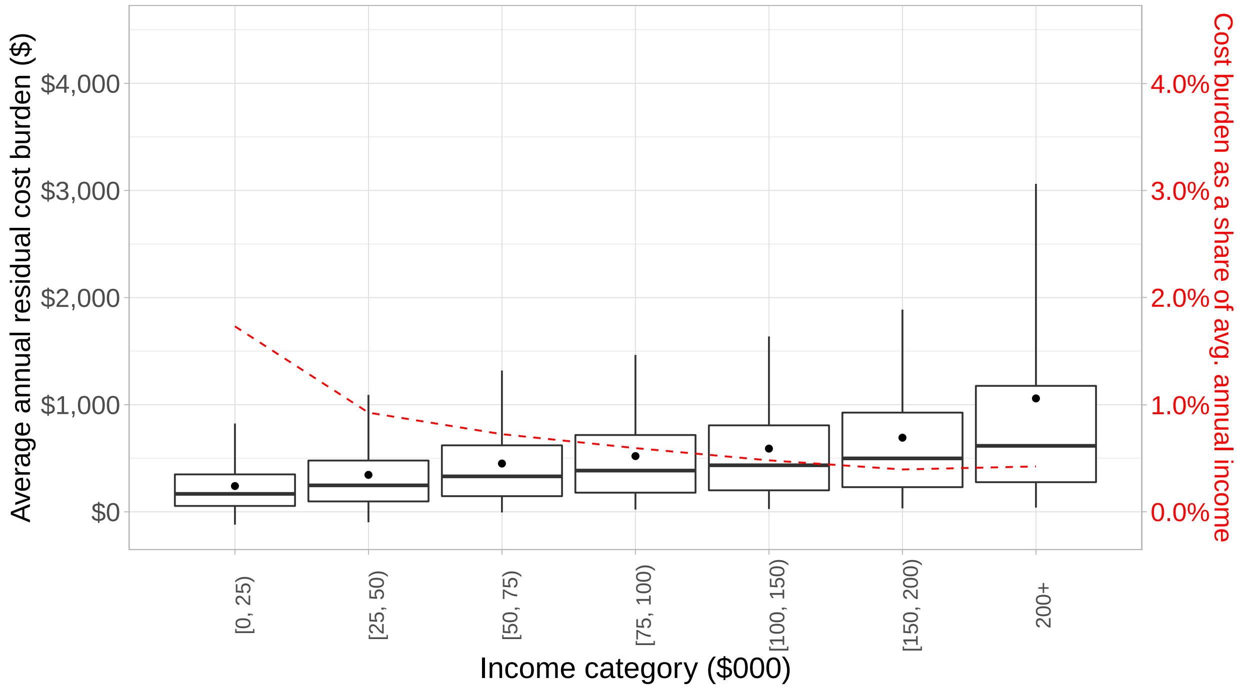


Figure 3: SCE Annual Residual Cost Burden (2019)
Annual bill increase from charging current rates rather than SMC

(a) $50 Social Cost of Carbon



(b) $100 Social Cost of Carbon

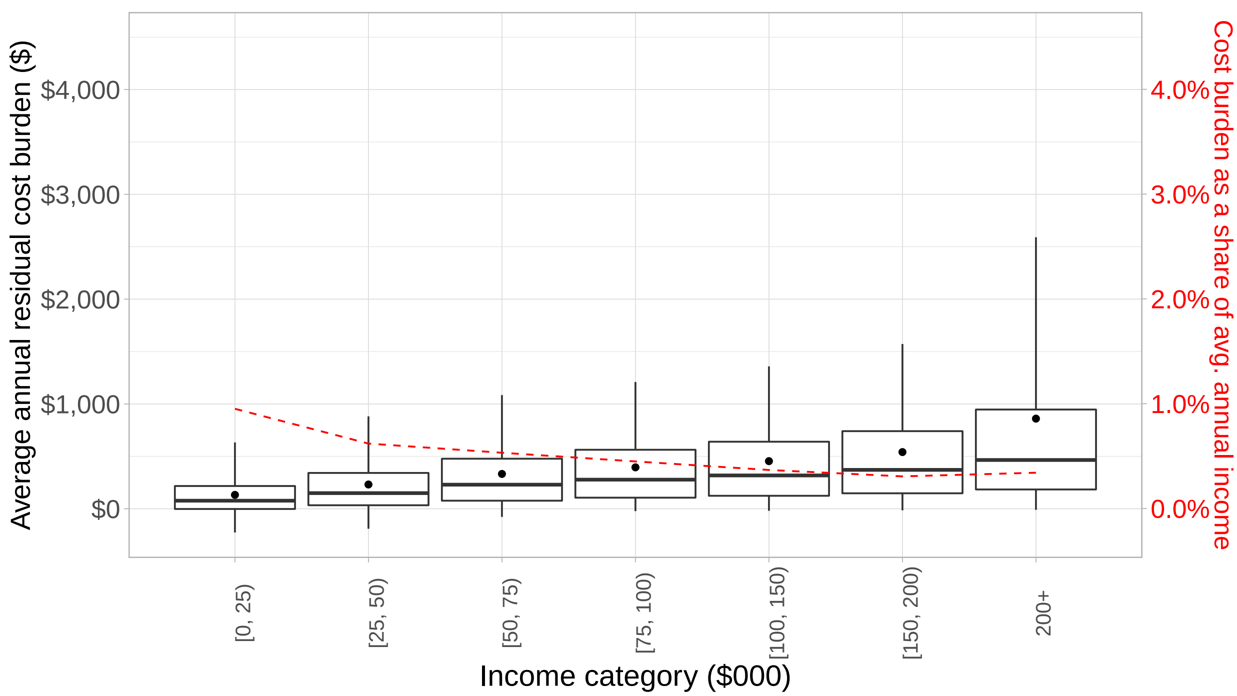
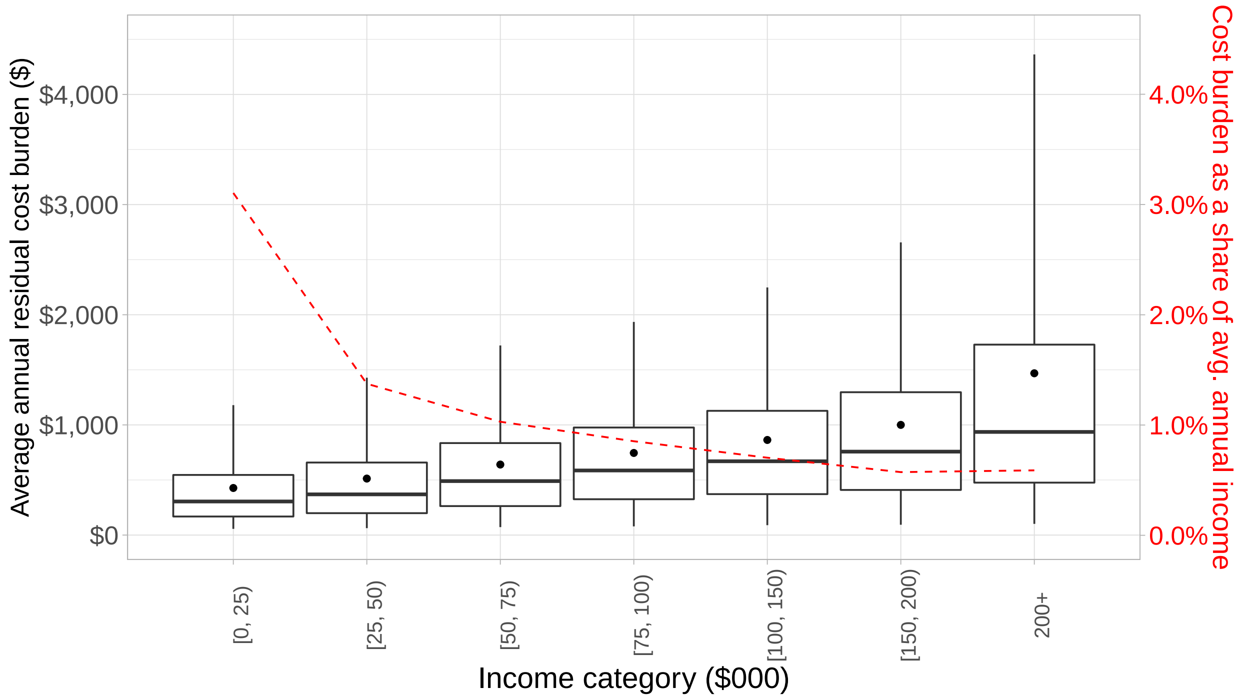


Figure 3: SDG&E Annual Residual Cost Burden (2019)
Annual bill increase from charging current rates rather than SMC

(a) $50 Social Cost of Carbon



(b) $100 Social Cost of Carbon

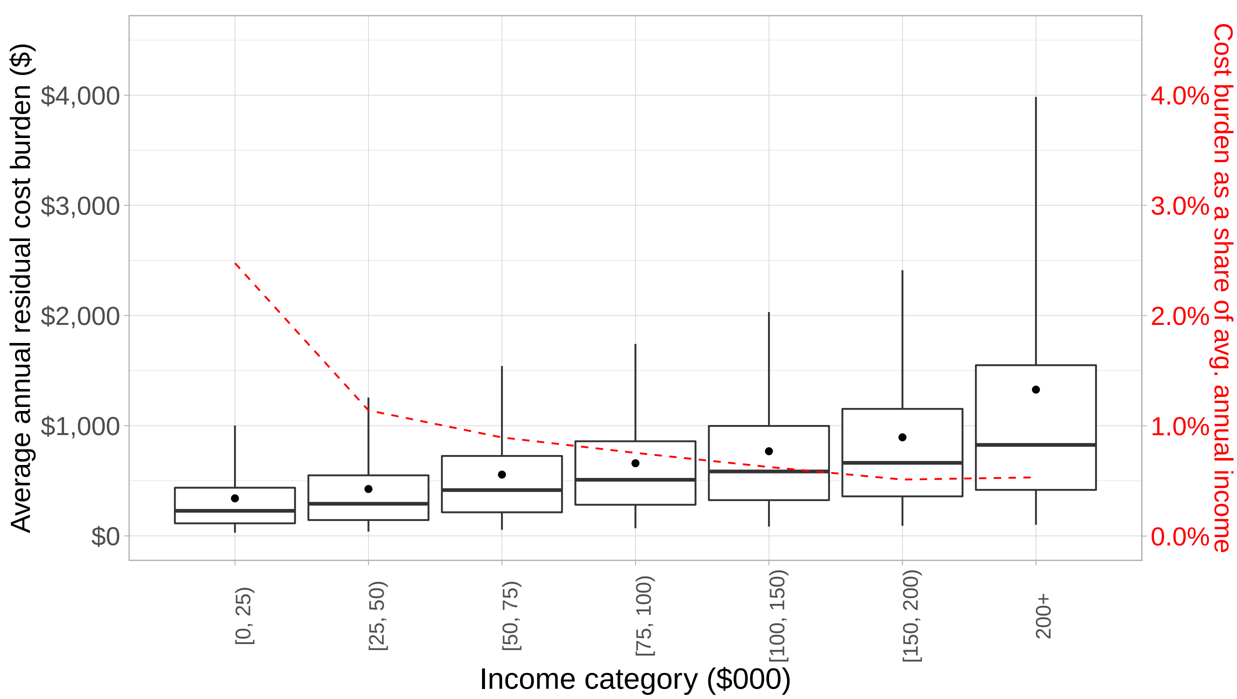
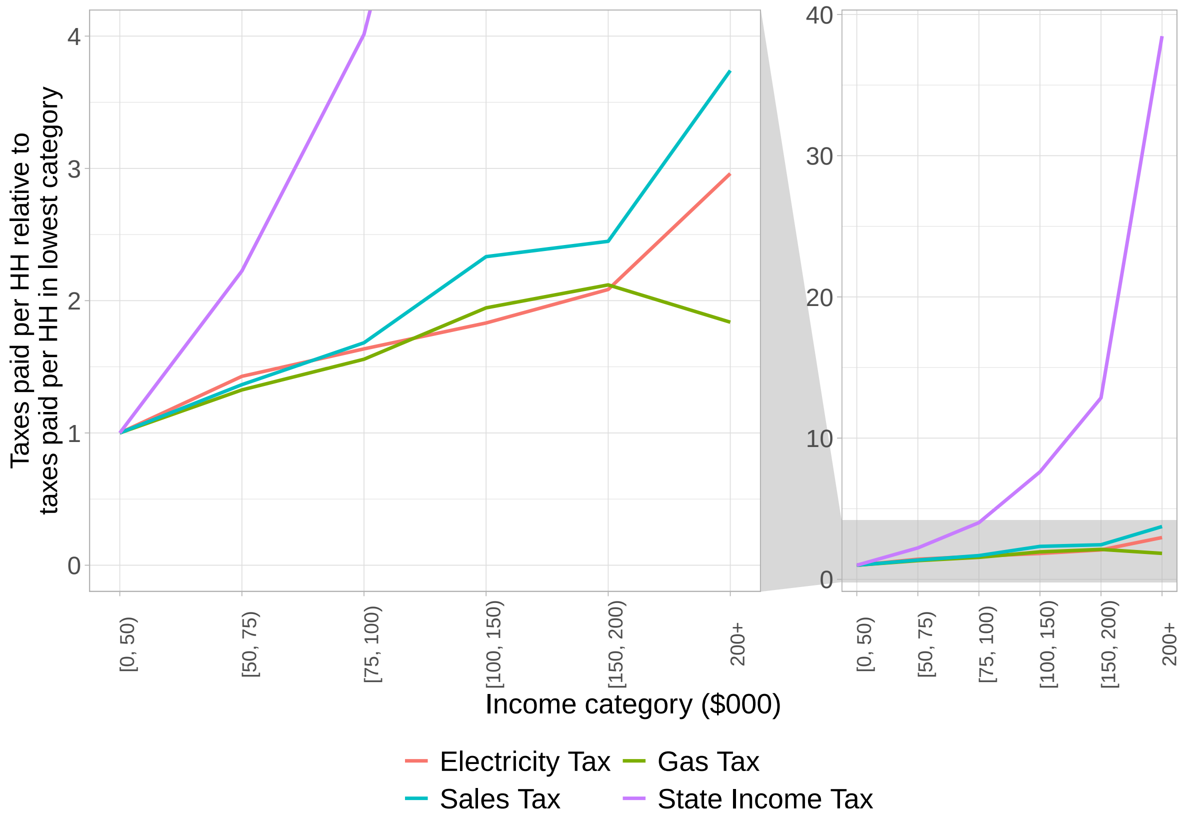


Figure 5: Progressivity of Different Taxes in California (2019)

(a) $50 Social Cost of Carbon



(b) $100 Social Cost of Carbon

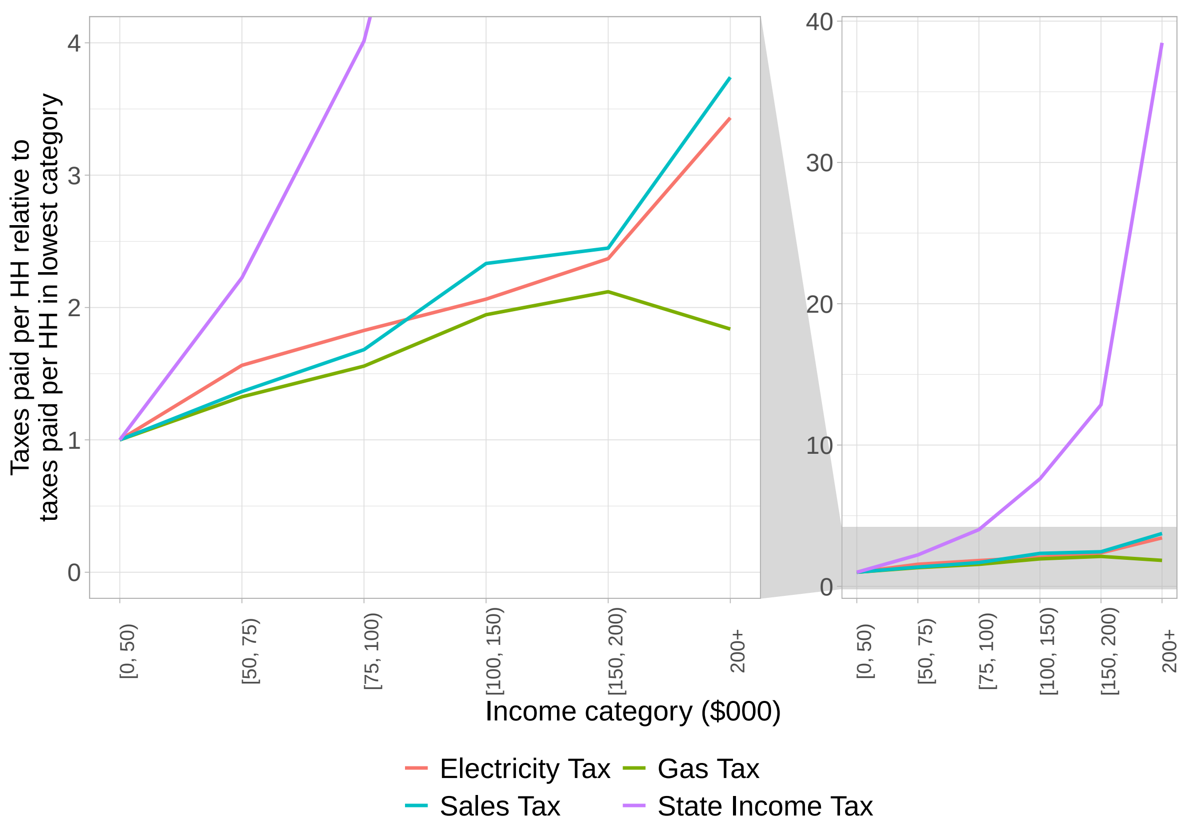
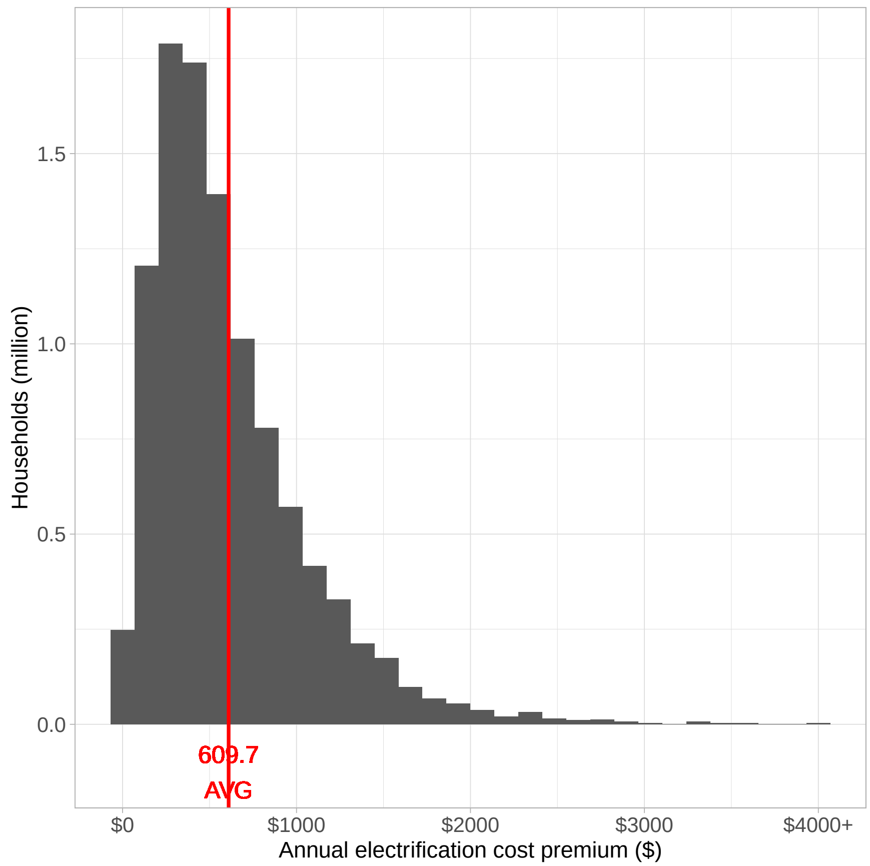


Figure 6: Distribution of Annual Electrification Cost Premia for Vehicles

(a) $50 Social Cost of Carbon



(b) $100 Social Cost of Carbon

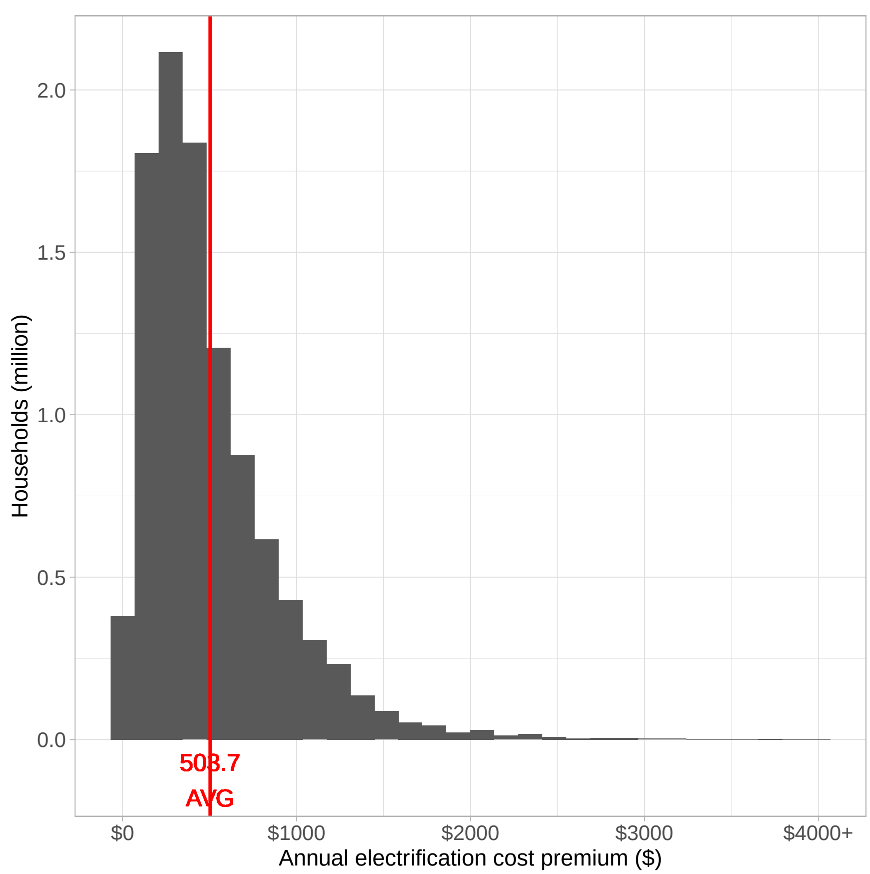
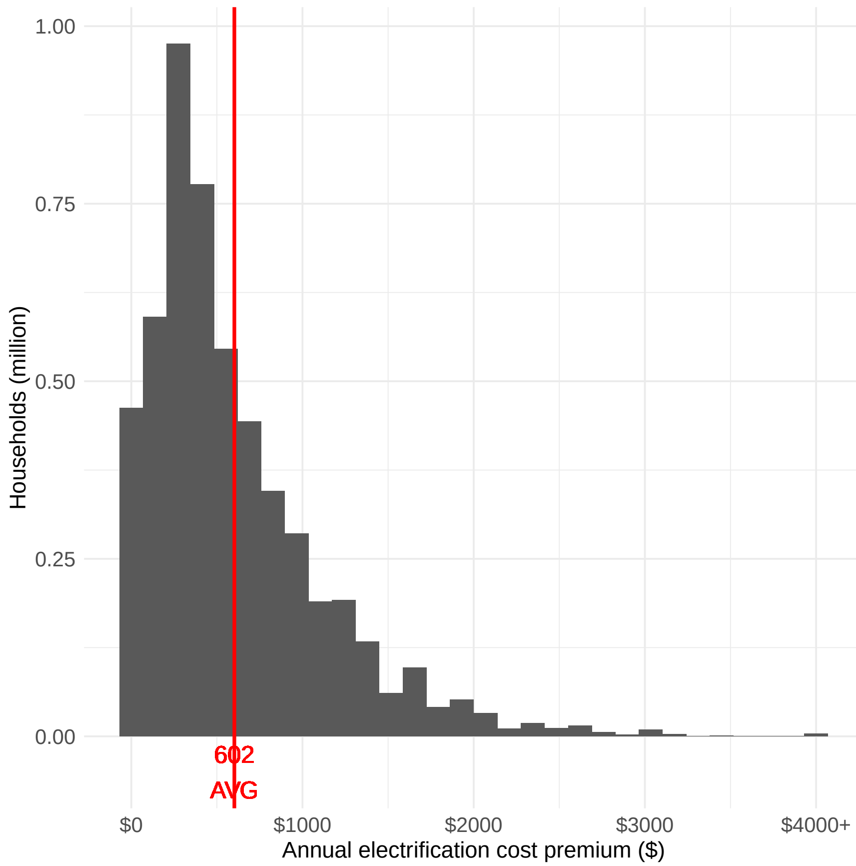


Figure 7: Distribution of Annual Electrification Cost Premia for Space Heating

(a) $50 Social Cost of Carbon



(b) $100 Social Cost of Carbon

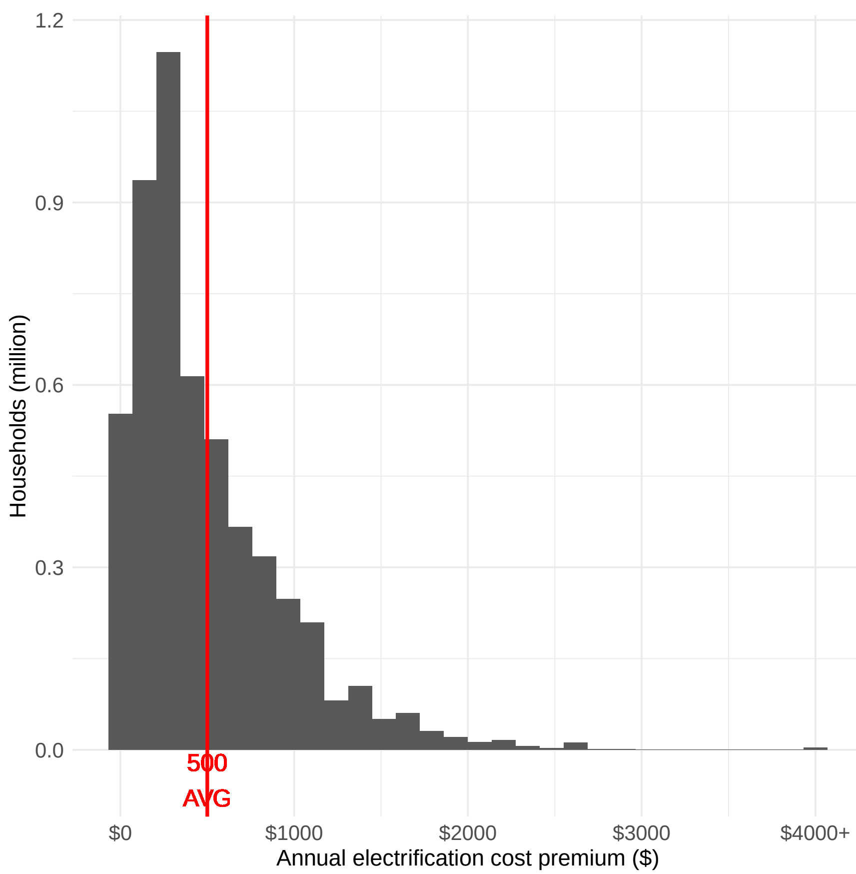
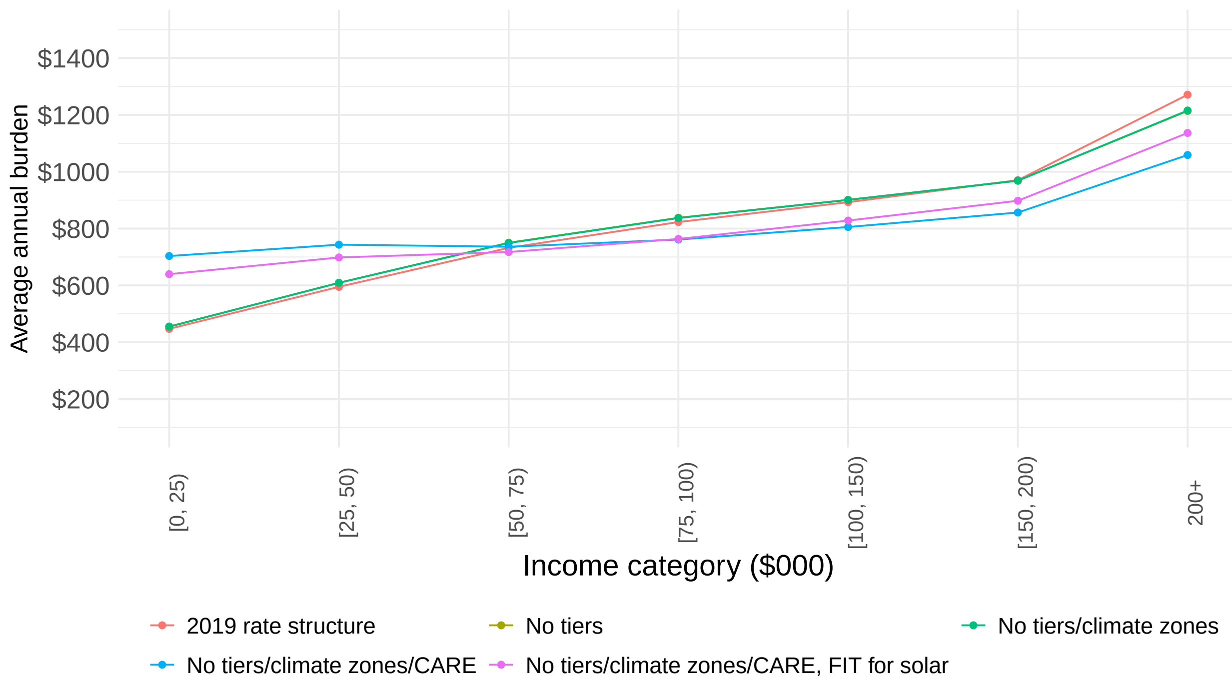


Figure 8: PG&E Annual Residual Cost Burdens Under Rate Alternatives (2019)

(a) $50 Social Cost of Carbon



(b) $100 Social Cost of Carbon

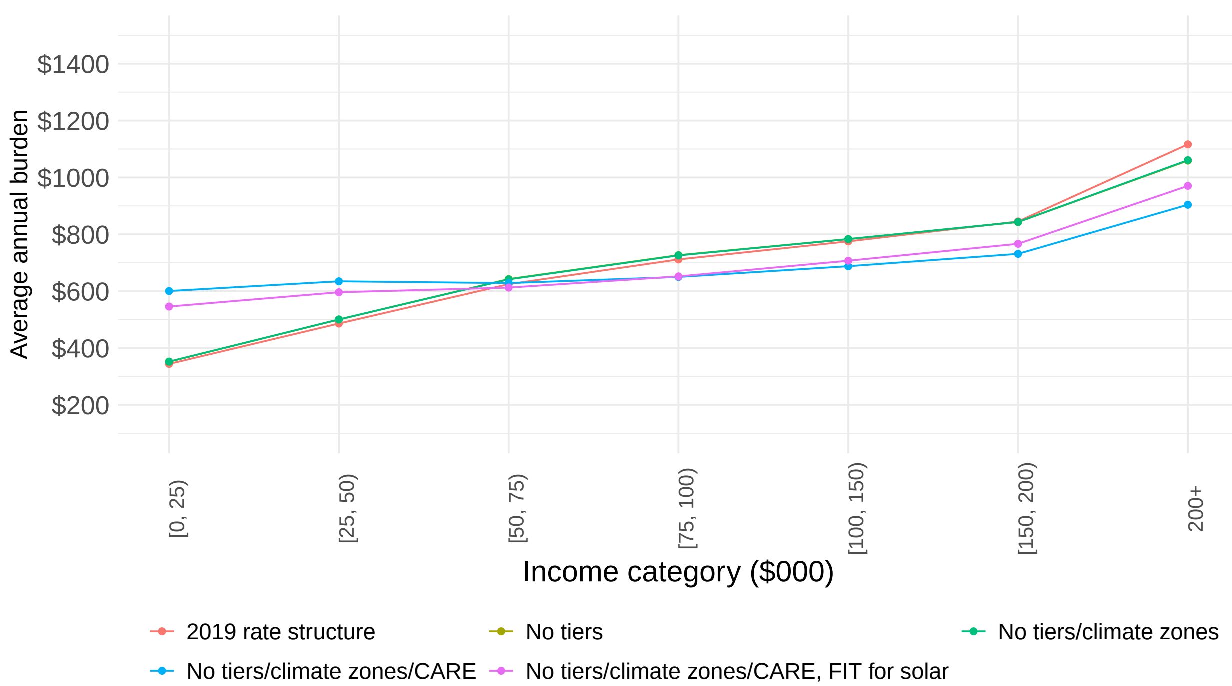
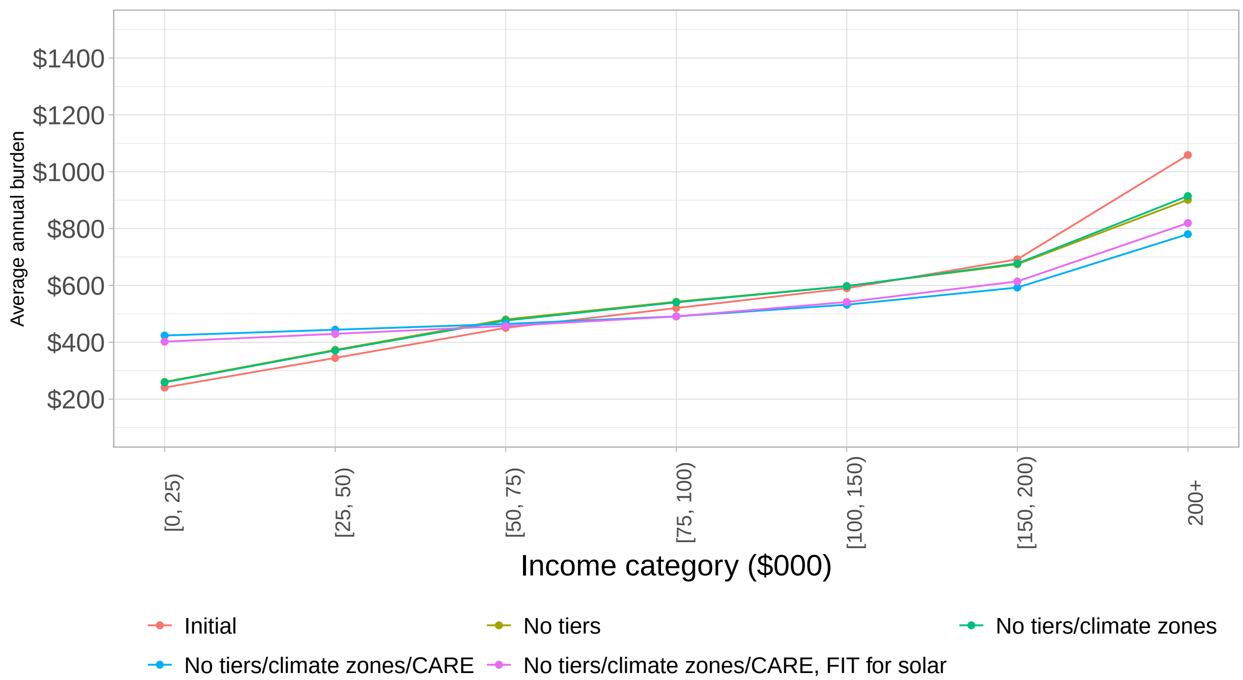


Figure 8: SCE Annual Residual Cost Burdens Under Rate Alternatives (2019)

(a) $50 Social Cost of Carbon



(b) $100 Social Cost of Carbon

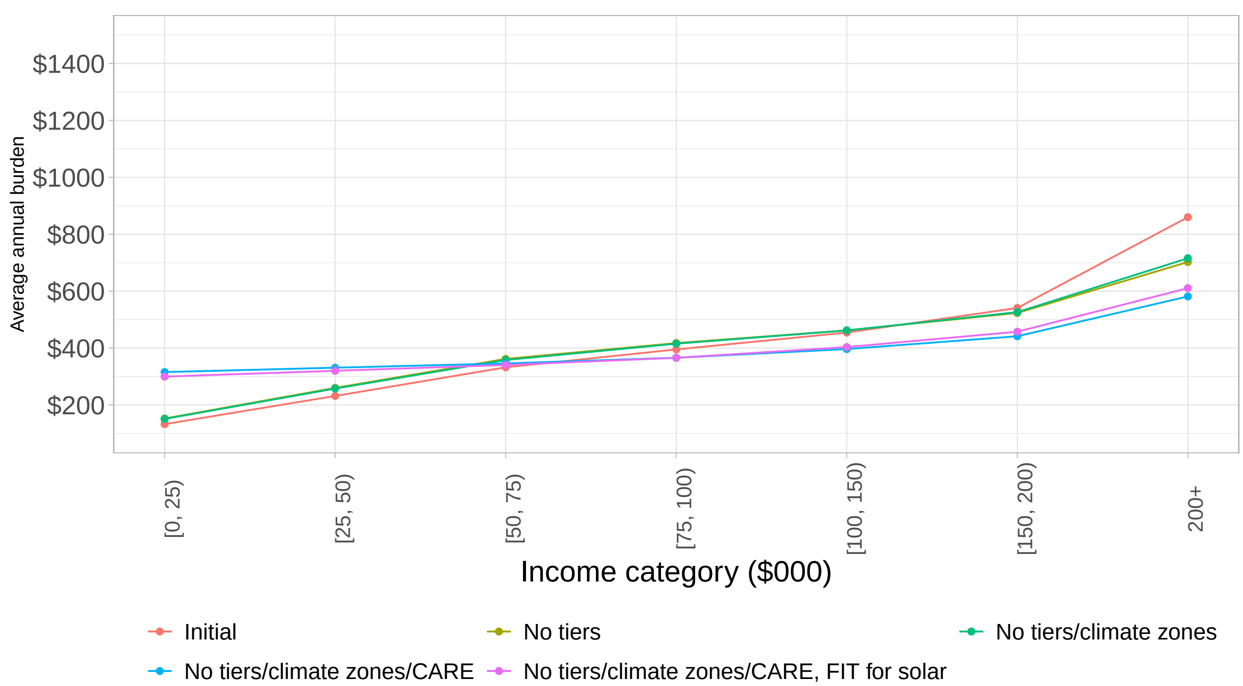
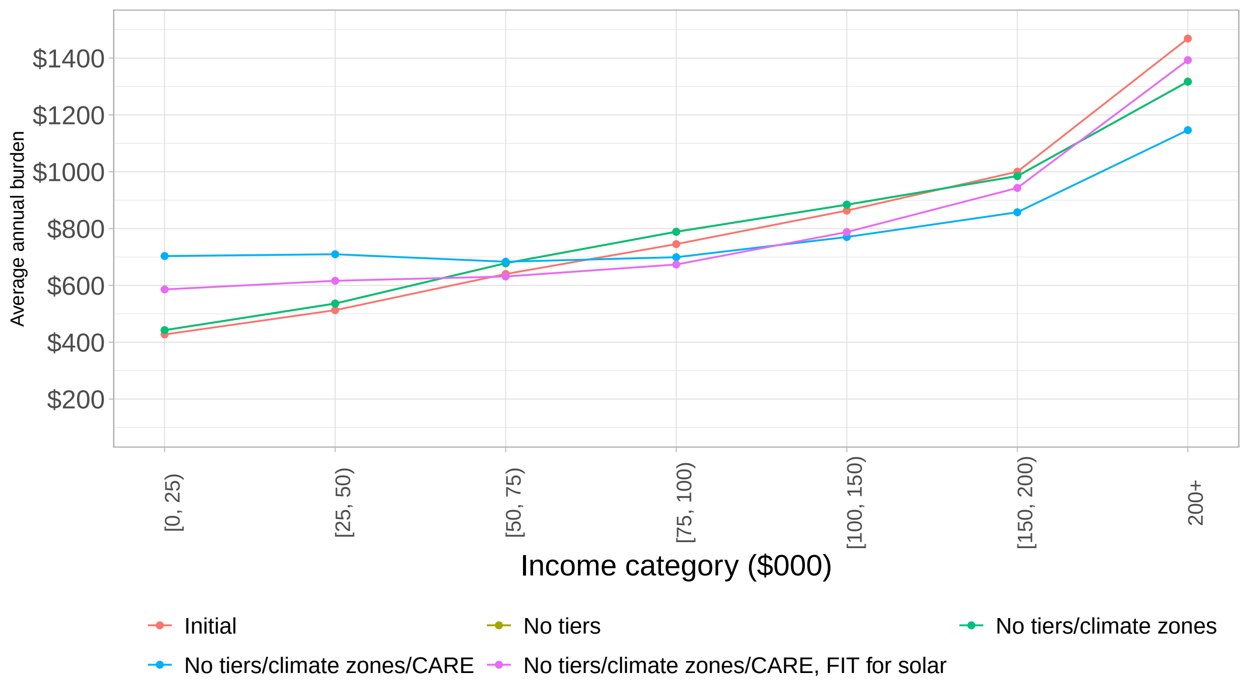


Figure 8: SDG&E Annual Residual Cost Burdens Under Rate Alternatives (2019)

(a) $50 Social Cost of Carbon



(b) $100 Social Cost of Carbon

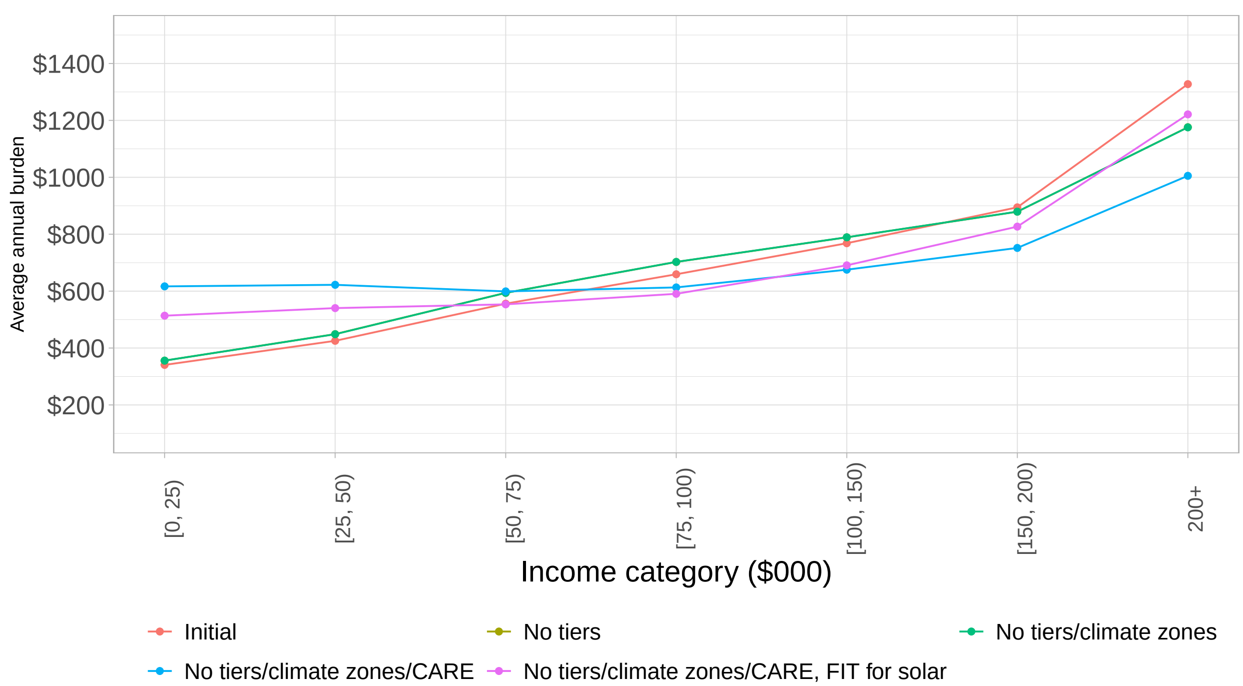
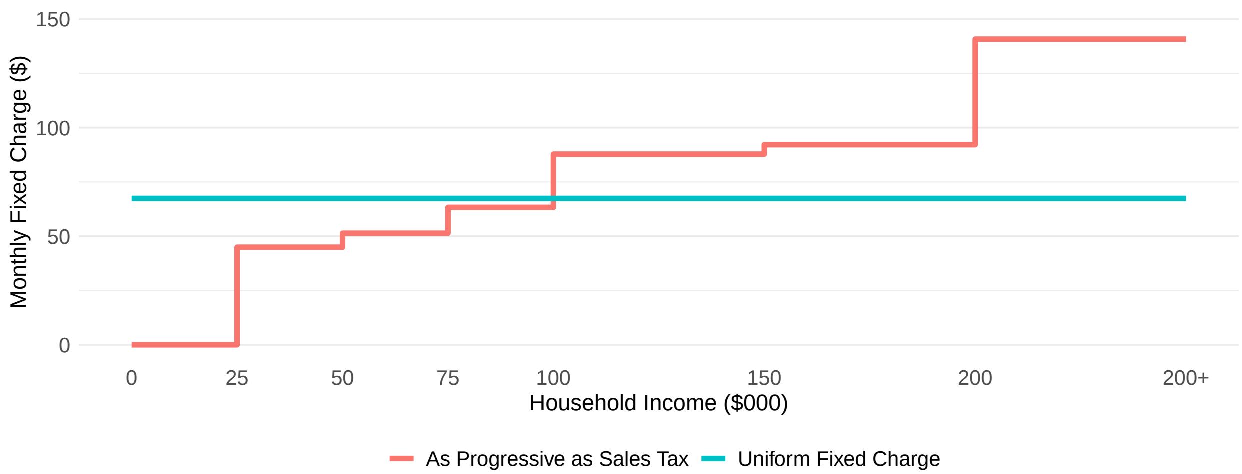


Figure 9: PG&E Monthly Fixed Charge Schedules

(a) $50 Social Cost of Carbon



(b) $100 Social Cost of Carbon

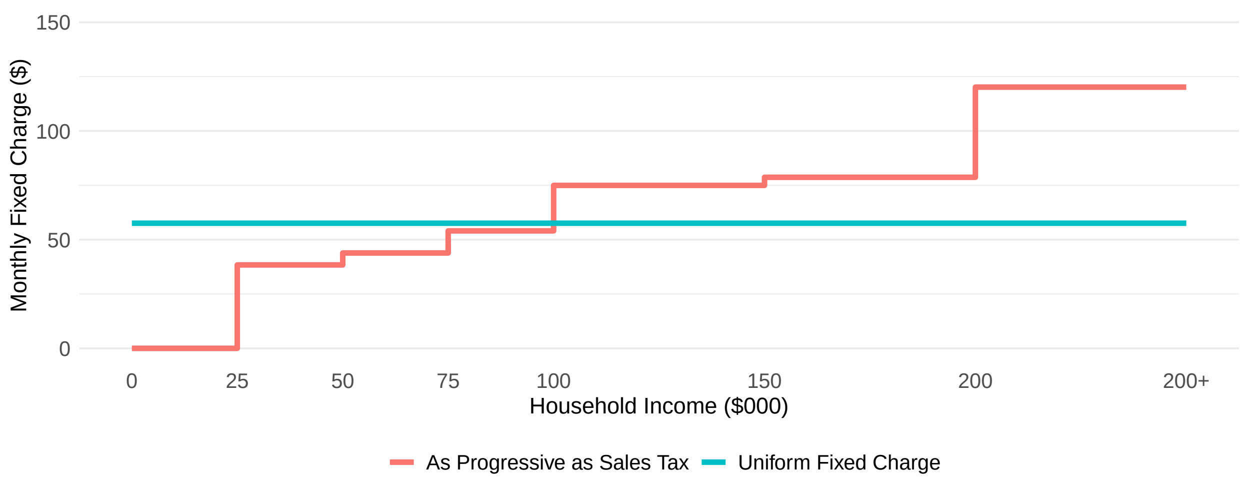
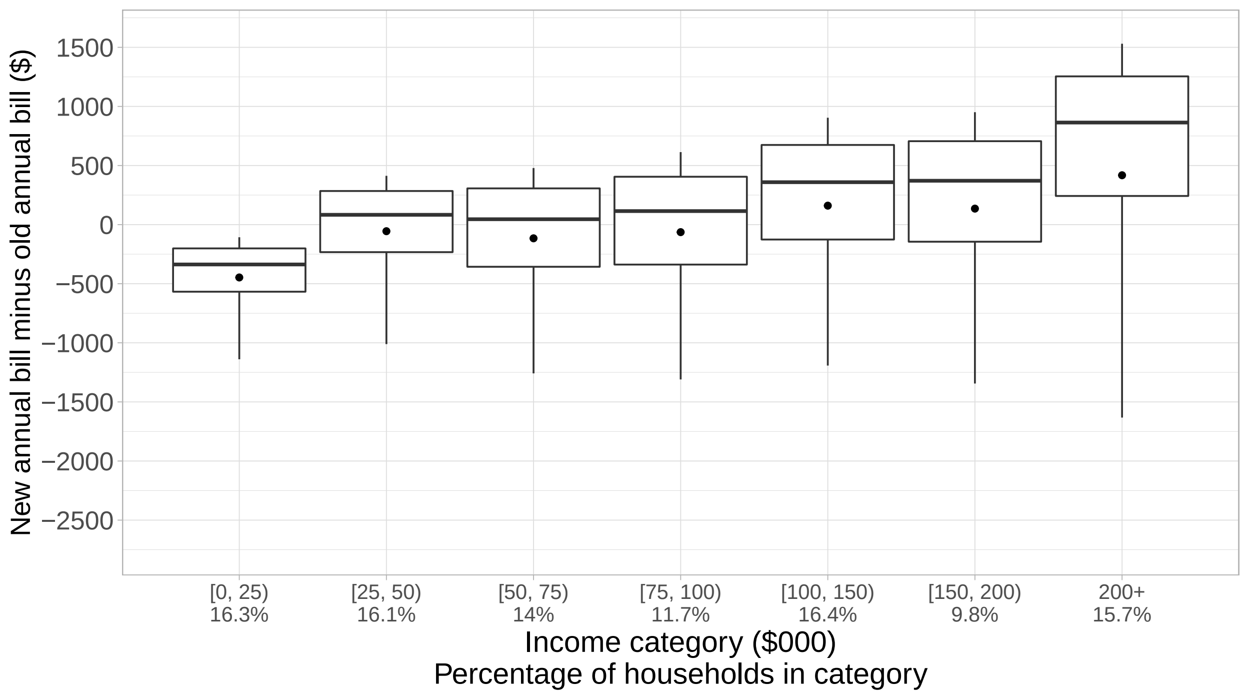


Figure 9: PG&E Change in Annual Bills Under Income-Based Fixed Charge

(a) $50 Social Cost of Carbon



(b) $100 Social Cost of Carbon

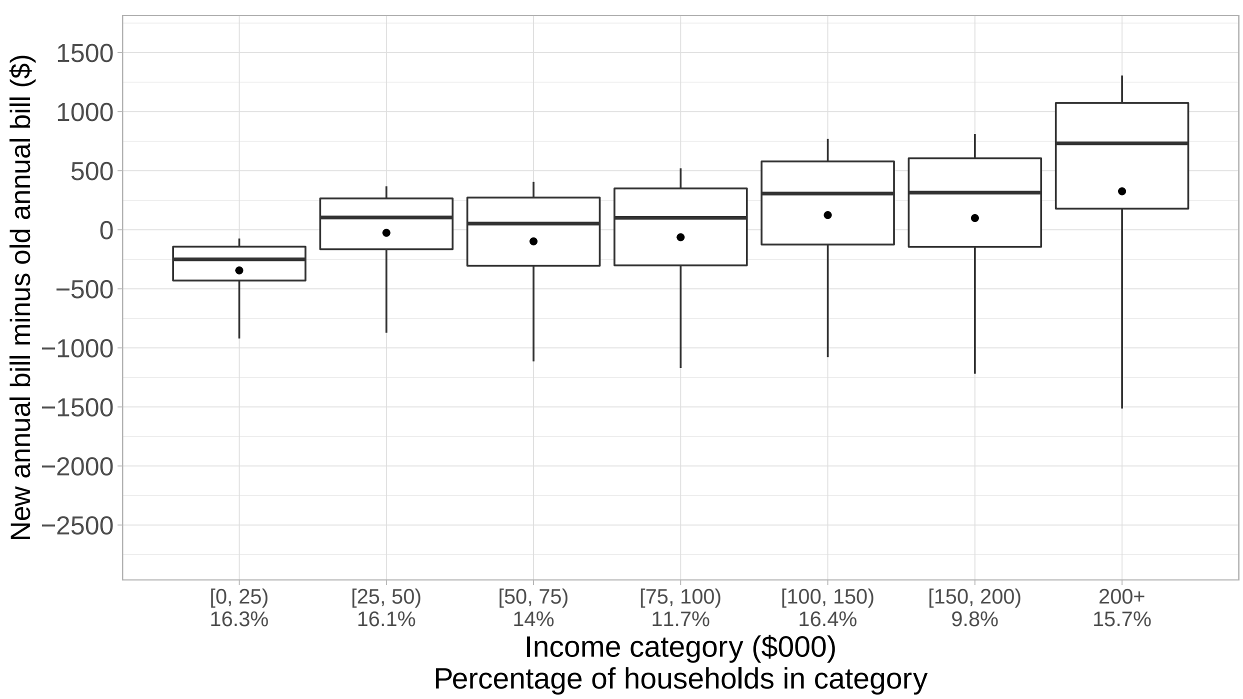


Figure 10: SCE Monthly Fixed Charge Schedules

(a) $50 Social Cost of Carbon

(b) $100 Social Cost of Carbon

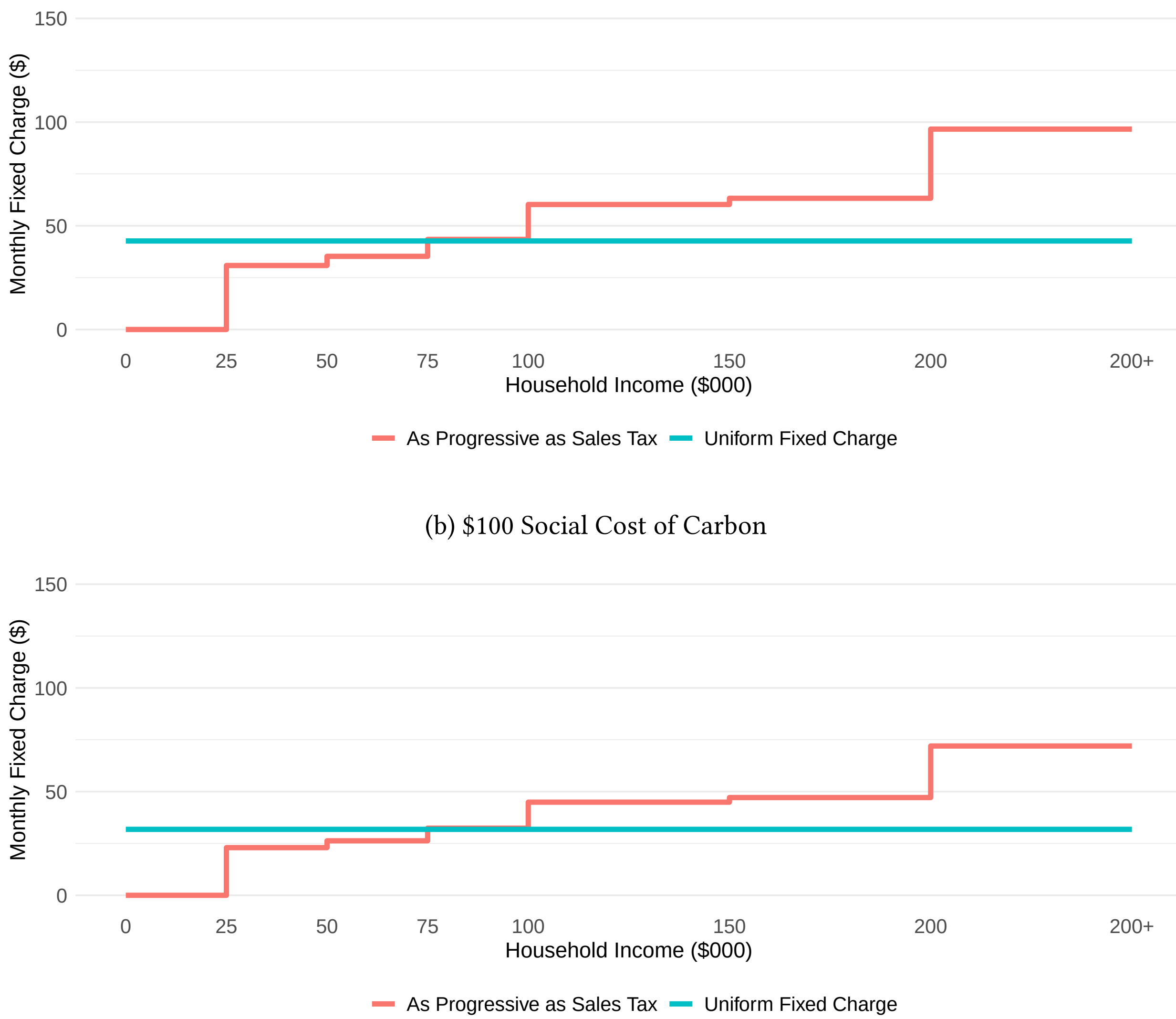
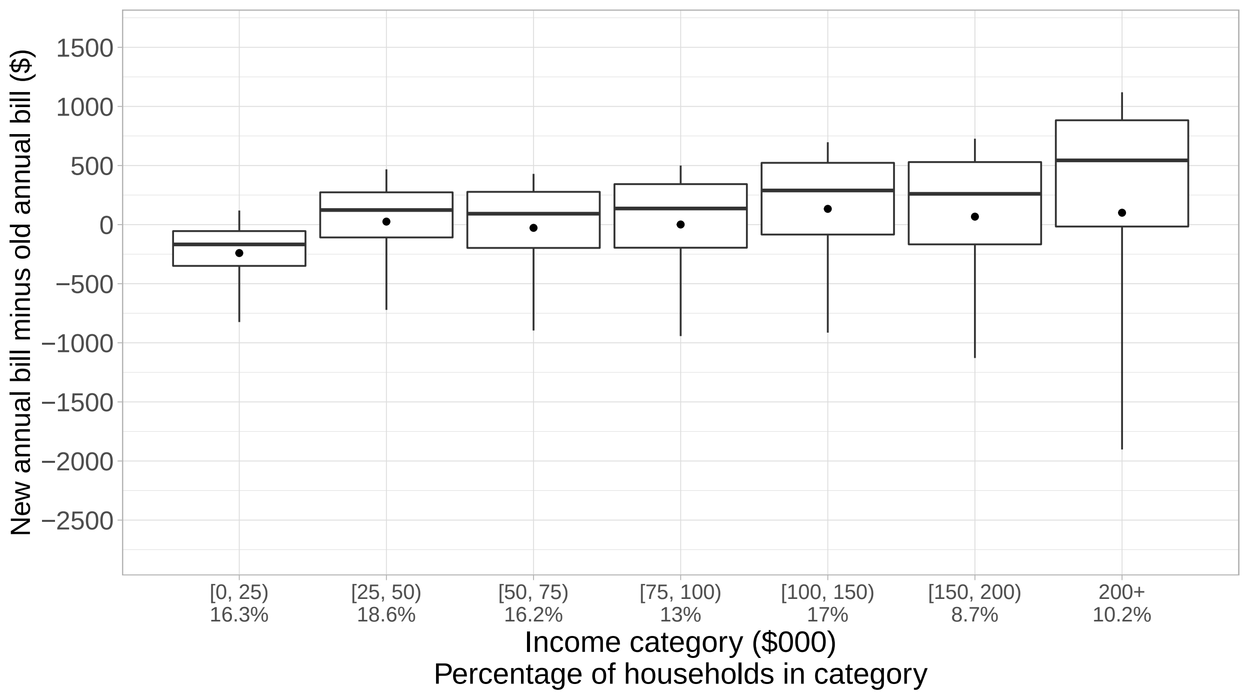


Figure 10: SCE Change in Annual Bills Under Income-Based Fixed Charge

(a) $50 Social Cost of Carbon



(b) $100 Social Cost of Carbon

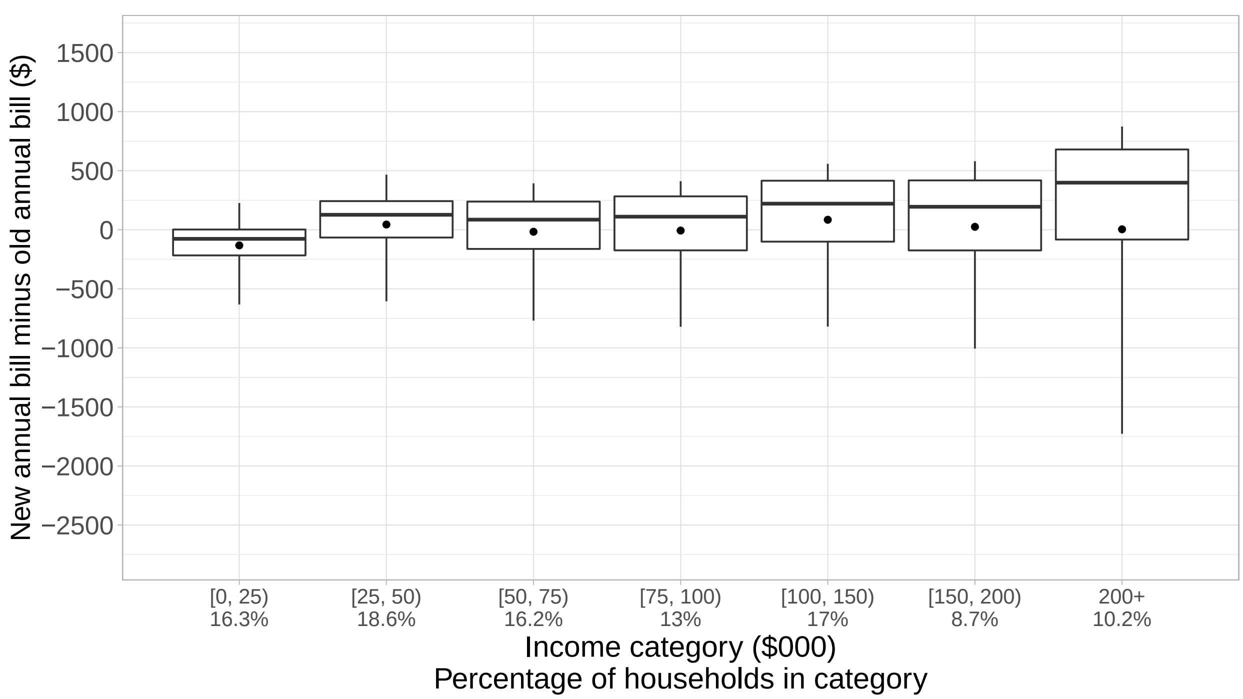
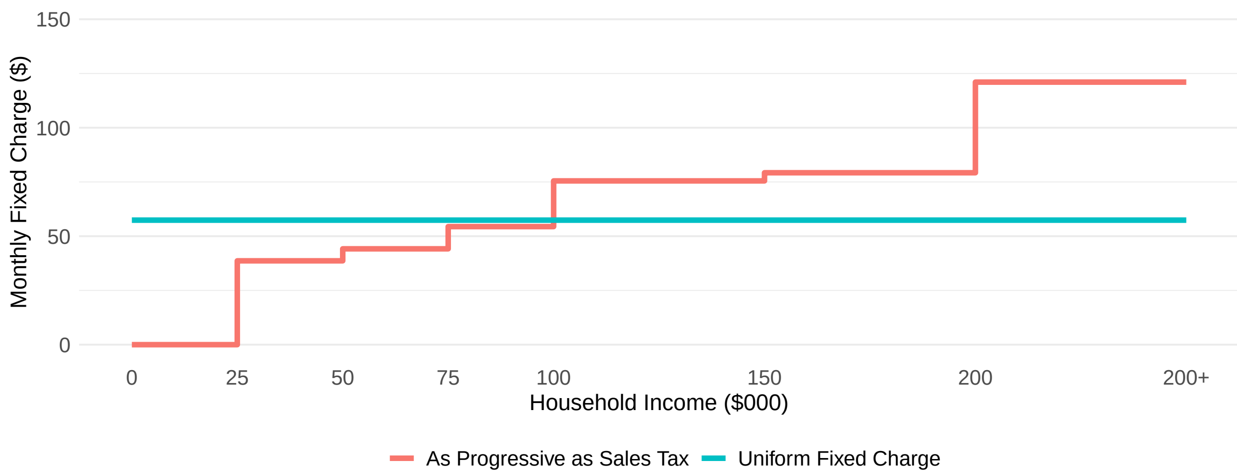


Figure 11: SDG&E Monthly Fixed Charge Schedules

(a) $50 Social Cost of Carbon



(b) $100 Social Cost of Carbon

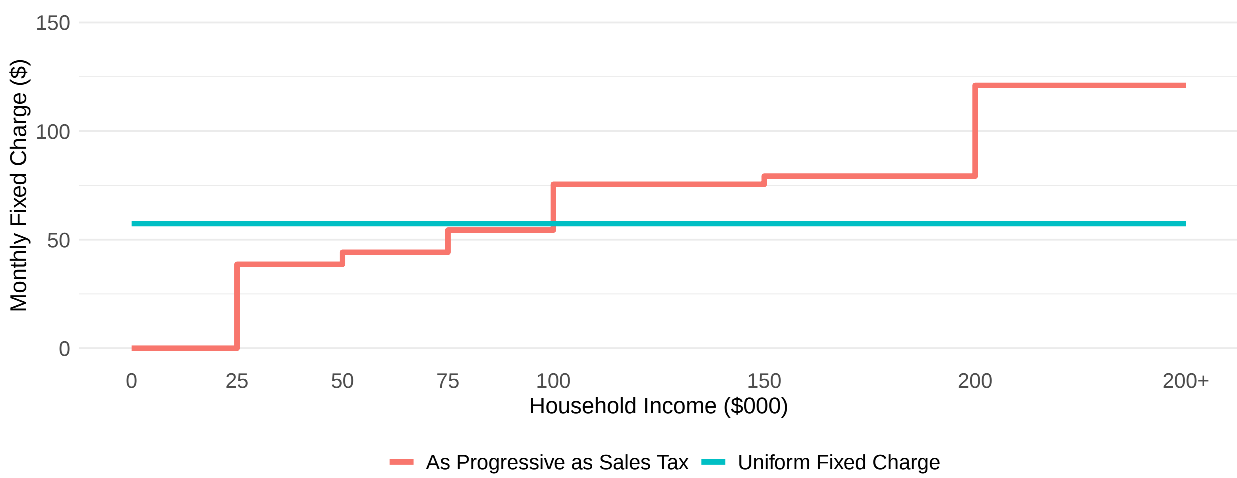
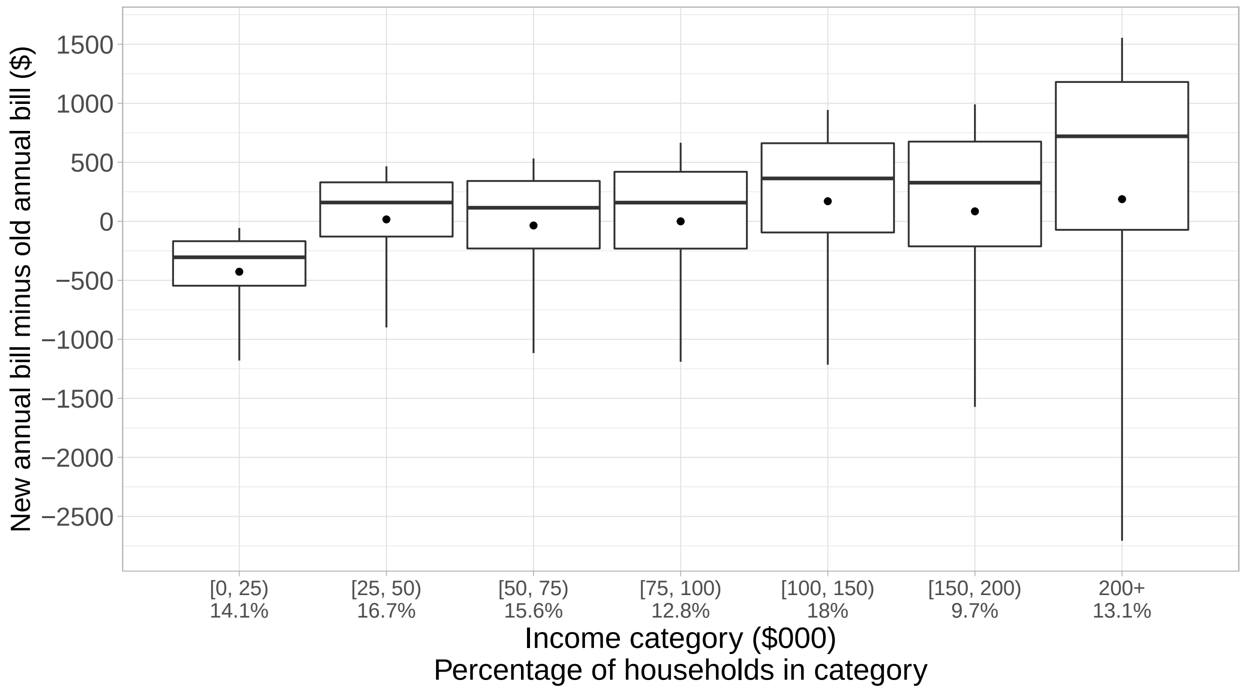


Figure 11: SDG&E Change in Annual Bills Under Income-Based Fixed Charge

(a) $50 Social Cost of Carbon



(b) $100 Social Cost of Carbon