An Aggregate Perspective on the Geo-spatial Distribution of Residential Solar Panels

Abajian, Alexander and Pretnar, Nick

University of California Santa Barbara

22 January 2021

An Aggregate Perspective on the Geo-spatial Distribution of Residential Solar Panels

Alexander Abajian[1] Nick Pretnar[2,3]*

[1]*University of California Santa Barbara*
[2]*University of California Santa Barbara, Laboratory for Aggregate Economics and Finance*
[3]*Carnegie Mellon University, Tepper School of Business*

January 22, 2021

Abstract

Residential solar panels in the United States (U.S.) are inefficiently distributed in terms of optimizing solar-electrical production. Controlling for local solar electricity generation potential (insolation), the residential solar share of electrical consumption is relatively higher in cloudier locales like the Pacific Northwest and Northeast than it is in sunnier areas like the Western U.S. and Florida. Rebates designed to increase residential solar adoption in places like Florida and Texas with relatively low solar-electrical shares are ineffective and may lead to net decreases in the residential solar share if housing and electrical consumption are complementary. This is because electrical consumption increases faster in response to a decline in effective residential solar prices than actual demand for panels themselves, thus driving down the solar share despite additional installations. Through the lens of a county-level structural model of demand for housing, electricity, and solar panels, we find that this phenomenon is especially prevalent in locales with high demand for cooling services (e.g., air conditioning, refrigeration, etc.) due to high numbers of cooling degree days. Inability to effectively store solar-produced electricity may be to blame. Our results thus suggest that future policies should subsidize nascent battery technologies in place of direct solar-panel installation rebates if the goal is to increase the residential solar share of electrical consumption.

Keywords: subsidies, environmental subsidy, environmental economics, electricity, energy utilities, renewable energy, solar energy, neighborhood characteristics, diffusion, spatial pricing, industrial geography

JEL Classification: H23, Q42, R23

*Corresponding author. UCSB, LAEF; 2112 North Hall, Santa Barbara, CA 93106-9215; npretnar@ucsb.edu.

1 Introduction

Uptake of solar photovoltaic (SPV) electricity generation technology among United States (U.S.) households has been strong over the past 20 years. The combination of favorable policy environments at the local, state, and federal levels and a 75 percent decline in the average cost (per Watt) of SPV systems since 2000 has driven the installation of almost 2 million residential solar systems through 2019 (Barbose et al. 2019). Distributed SPV, a broader measure capturing all non-utility scale installations, comprised 4.5 percent of U.S. electric generating capacity in 2018, while its share of total generation will likely surpass 1 percent in 2020 (EIA 2020).[1]

However, new residential installations are distributed unevenly and inefficiently around the country. Cloudy, cool places receive lower influx of solar radiation that is convertible to electricity, yet many of these places, like the Pacific Northwest and Northeast, have seen relatively high levels of residential solar adoption compared to sunnier and warmer environments like those in Texas and Florida. Further, we observe such disparities in residential solar adoption despite the fact that rebates and grants to recover the cost of installation are highest in both Texas and Florida. This fact leads to several questions which we explore in this paper. Do political-party preferences play a role in residential solar adoption, above and beyond local fiscal incentives? Are climatological factors, beyond relative solar insolation from high rates of sunny days, at play which lead to conditionally higher adoption rates in colder environments? Are rebates inefficiently deployed given local political preferences or climatological factors weighing on the residential solar share of electrical consumption? If so, can we construct incentives that lead to both aggregate increases in the solar share, and a more geo-spatially efficient solar panel distribution which takes advantage of the implicit un-tapped energy in sunnier locales?

We explore these questions by combining several different datasets that record local costs and incentives associated with solar panel adoption, as well as various demographic, cultural, and climatological characteristics of U.S. counties. Compared to what we would expect given the potential for electricity generation from solar panels, demand for residential solar in sunnier places is relatively lower than in generally cloudier environments when we account for differentials in the potential for solar-energy production (insolation). This is true even for sunny areas with high levels of rebates and incentives to encourage adoption. We also find that the number of cooling degree days (days for which air conditioning is required to maintain a specific building temperature) negatively and

[1] We use the terms distributed and small-scale interchangeably to denote all non-utility scale SPV generation.

significantly depresses the local residential solar share of electrical consumption, helping explain the relative inefficiency of the aggregated geo-spatial distribution of residential solar across the U.S.

Given these empirical facts, we use a structural approach to show that subsidies and rebates alone likely cannot offset the negative impact of demand for cooling services, namely from air conditioning, on the solar share of electrical consumption. Absent the availability of storage technology, households in sunnier, warmer areas with residential solar may place a strain on existing grid infrastructure when they experience lapses in insolation. During periods of low insolation, lack of storage precludes households from using the excess electricity they produce in periods of high insolation. This could lead to a net decline in the solar share of total consumption if households that install residential solar also increase total electrical consumption on the margin, which our findings suggest. It is likely that storage technology could ameliorate the downward pressure that demand for cooling services places on the solar share. Thus, policies that subsidize and incentivize both residential solar adoption and the installation of batteries for storage could be one mechanism to help achieve permanent increases in the solar share and possibly lead to a more efficient aggregate, geo-spatial distribution of panels.

This paper proceeds as follows. In Section 2 we provide context for the questions we explore here. In Section 3 we present model-free evidence that the aggregate, nation-wide geo-spatial distribution of solar panels fails to efficiently take advantage of insolation in sunny locales, and that panel-installation rebates likely are not positively contributing to geo-spatial efficiency. Section 4 describes a structural model of household demand for housing and electrical consumption that accounts for heterogeneity in local, solar amenity preferences. Section 5 discusses the structural estimation procedure and corresponding parameter estimates. Section 6 contains counterfactual experiments designed to understand which factors most significantly affect the residential solar share of electrical consumption and the geo-spatial distribution of solar panels. Finally, Section 7 concludes.

2 Background

Optimal adaptation and allocation of distributed solar generation is given substantial attention in the academic literature. The natural sciences have covered a myriad of technical aspects ranging from developing user tools for optimal panel placement to forecasting regional production and storage potential (Freeman et al. 2018; Sengupta et al. 2018). Economists have investigated the direct impacts of policies intended to encourage renewable energy adoption. Public research and development (R&D) expenditure (Wiser and

Millstein 2020), state and local incentives (Hughes and Podolefsky 2015), and Federal tax expenditure (Borenstein and Davis 2016) have been scrutinized due to the contentious nature of public discourse on energy policy.

Other strands of the economic literature pose important questions about how to incorporate new renewable capacity into the electrical grid. Empirical linkages between utilities' transmission price schedules and adoption of distributed solar capacity have been established at the local level (Borenstein 2017; Wolak 2018). The price effects of zero marginal cost generation and the large fixed costs associated with solar installation have raised concerns about optimal rate setting for both a future fully-renewable grid and along transition paths toward such a future (Imelda, Fripp, and Roberts 2018; Heal 2020).

Further studies have examined the welfare effects of solar adoption through the channel of pollution mitigation. Researchers have used time series data on power sector pollution to quantify the degree to which social costs stemming from emissions displaced by increased renewable generation have been avoided. The localized environmental benefits are highly dependent on where new renewable capacity is installed and what form of non-renewable electricity generation is displaced (Callaway, Fowlie, and McCormick 2018; Sexton et al. 2018). Carbon emissions' abatement costs associated with subsidizing the diffusion of small and utility-scale solar distribution have been studied by both academic and government bodies, but findings are wildly mixed in terms of abatement cost estimates (see Gillingham and Stock 2018 for a recent survey).

Additional research has gone towards better-understanding renewable energy uptake due to differences in regional preferences over electricity generation. Nomura and Akai (2004) and Yoo and Kwak (2009) find heterogeneity in preferences for renewable electricity in general, as well as heterogeneity in preferences for different types of renewable sources. Heng et al. (2020) and Bakkensen and Schuler (2020) present survey evidence of premia in individuals' willingness-to-pay for renewables and its sensitivity to variations in energy policy. Bollinger and Gillingham (2012) find that peer effects are significant in determining residential solar installation patterns: a consumer is more likely to install residential solar panels if his neighbors have also installed them.

Building on the evidence presented in the literature, we construct a micro-founded equilibrium model of residential solar panel demand that directly incorporates heterogeneous amenity preferences for local green energy generation. We combine datasets from Stanford's DeepSolar project and the Lawrence Berkeley National Lab (LBNL) to form county-level estimates of residential solar generation, insolation (the amount of solar radiation reaching a given area), and the lifecycle costs of residential solar electric generation. We use these data to structurally estimate aggregate preference parameters associated with housing, electrical, and the demand for residential solar panels at the U.S. county level. This choice of model specification allows us to analyze how heterogeneity in climatic factors and solar subsidization policies across localities affects the aggregate geo-spatial distribution of residential solar electrical consumption across the U.S. Through the lens of the model, we can then explore whether certain subsidization policies could lead to a more efficient geo-spatial distribution of residential solar panels, so that the solar share of electrical consumption is highest in sunny places where solar electricity can be most feasibly generated.

3 A Snapshot of U.S. Residential Solar

3.1 Geographical Panel Coverage

The most granular spatial data on solar panel coverage come from Stanford University's DeepSolar Project measuring solar panel deployment across the United States (Yu et al. 2018). The project utilizes machine learning techniques to analyze satellite imagery and infer the location and surface area of small-scale solar panel installations across the lower-48 U.S. states. The data is currently available as a static snapshot that was last updated in December 2018. This dataset links solar panel location and surface area with average daily solar insolation, measured in kilowatt-hours per square meter per day ($kWh/m^2/d$), at the Census-tract level. Our goal here is to utilize this unique dataset to first get a general picture of how solar panels are distributed across the U.S. We would like to understand the degree to which the current deployment of solar panels is geo-spatially efficient, and to understand how different policies, like rebates and grants, affect the efficient distribution.

For a cursory understanding of how solar panels could be most-efficiently deployed, Figure 1 shows a choropleth map of county-level average daily solar insolation in units of $kWh/m^2/d$. This measure captures the maximum amount of energy a square meter of SPV cells could produce over the course of the average day in a given location.[2][3] If electricity were both storable, in say large batteries, and could freely flow across a national network, then *optimal* solar panel deployment would follow the gradient featured in the

[2]The median and mean for the 3,097 U.S. counties in the lower-48 states with available data are 4.12 $kWh/m^2/d$ and 4.18 $kWh/m^2/d$ respectively. Note that 19 of the 3,116 counties and county-equivalents in the lower-48 states lacked insolation data at the census-tract level.

[3]Mason County in the state of Washington has the lowest average daily solar insolation at 3.35 $kWh/m^2/d$, while Santa Cruz County in the state of New Mexico features 5.55 $kWh/m^2/d$.

choropleth map in Figure 1.[4] Under optimal allocation we would have high concentrations of panels on land in the highly-irradiated Southwest and low concentrations in the Pacific Northwest and Northeast. The electrical grid would then be structured so as to transport this electricity from high-insolation to high-consumption locations. However, physical restrictions (among other issues) cap cross-country transmission capacity, which is one contributor to rich variation in the distribution of residential installations.

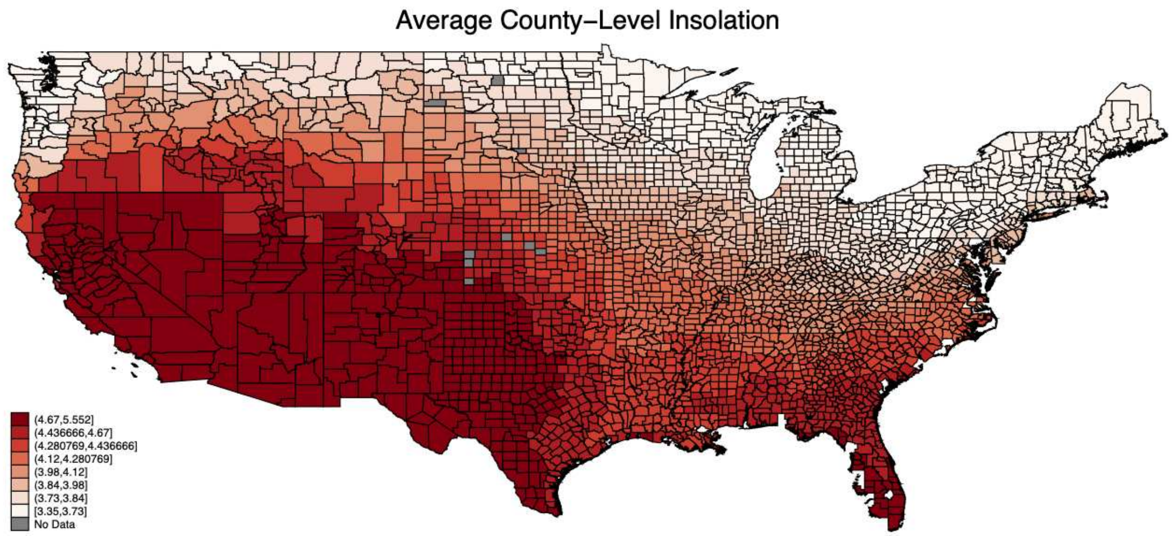


Figure 1: In this figure we present county-level average daily solar insolation ($kWh/m^2/d$) where the colors represent octiles, with the brighter reds corresponding to places on the map with high solar insolation. Note that the Western U.S. and Florida have climatological conditions which contain the most relative solar-energy potential, while the Pacific Northwest and Northeast feature the least relative solar-energy potential.

Figure 2 displays the rich dispersion in penetration of residential solar installations.[5] The color gradient in this figure is divided into octiles of percentages of households at the county level that have installed solar capacity (as of December 2018). While solar panel uptake is highly concentrated in the Southwest wherein some counties have uptake rates of over 10 percent, household installations are prevalent along the Eastern seaboard as well. Despite the diffuse nature of panel uptake, daily-insolation and aggregate residential solar capacity are significantly correlated at the county level.

[4]Note that optimal allocation refers only to efficiency in generation per area of SPV cells, abstracting (for the moment) from variation in installation costs.

[5]We should note that due to lack of data on installation dates or costs in the DeepSolar dataset, for more detailed analyses we must turn to the LBNL data which features installation data for only a subset of American counties. Our structural model will thus ultimately operate on a smaller geographic sample than what is presented in Figure 2.

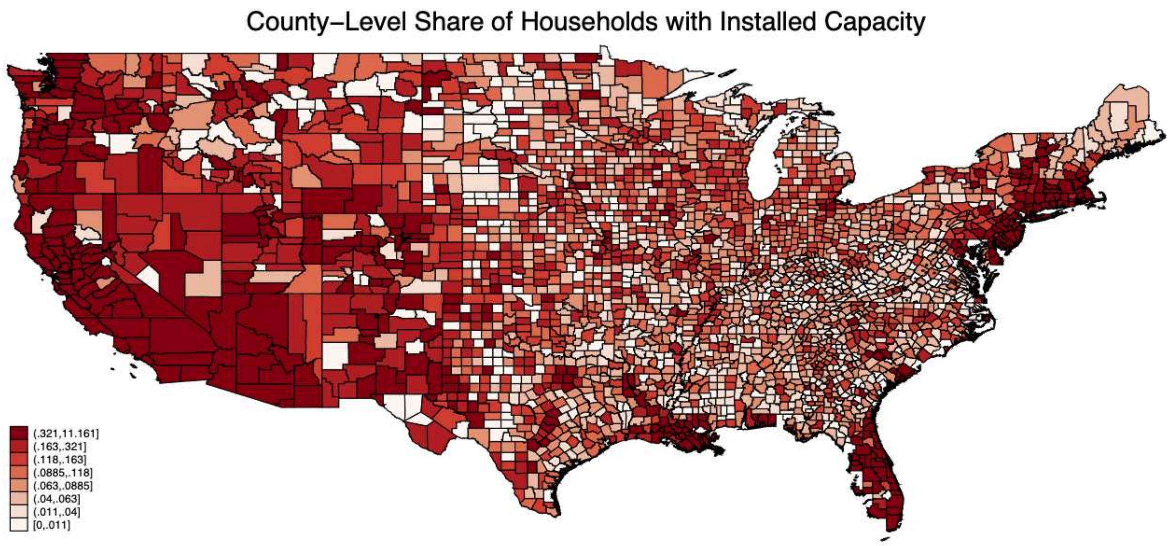


Figure 2: In this figure we present the percentage of households in a given county that feature residential solar installations. Again, the colors represent binned octiles, with brighter reds corresponding to places with greater residential-solar uptake.

3.2 Estimating Residential Solar Generation

Unfortunately, DeepSolar, LBNL data, and administrative sources all lack direct measurements of residential solar electricity generation and consumption at the county level. The Energy Information Administration (EIA) estimates small-scale solar generating capacity and generation at the state level using respondents' annual estimates in form EIA-861.[6] The EIA approach relies on utilities' reporting of net-metered capacity along with distributed and dispersed resources that are not net metered. We reconcile the county-level estimates for generation that can be inferred from the LBNL and DeepSolar data with the state-level estimates from the EIA and construct an adjusted series at the county level.

While DeepSolar provides estimates of panel area by household, the LBNL dataset contains actual records of residential solar installations. Specifically, LBNL catalogs 1,456,836 observations of residential solar installations between 1998 and 2018. Among the residential observations, 1,021,967 are complete observations containing all information necessary to calculate the value of residential solar electricity. Observations from the states of Utah, Maryland, Rhode Island, and Colorado are dropped, as none contain zip code level location data. The share of residential observations missing zip codes is negligible in other states covered in the dataset.[7] Dropping other incomplete observations removes

[6]For more details visit: https://www.eia.gov/electricity/monthly/pdf/technotes.pdf
[7]Entries missing zip codes comprise less than 1% of total residential system capacity in all other states.

an additional 83,799 data points from the dataset.[8]

To calculate average household solar generation at the county level, we use the LBNL data, which contains records of both panel size and efficiency rating. Let i denote individual observations in a given county. Let χ_{ics} denote system-level efficiency and $\bar{\phi}_{cs}$ denote average daily insolation in county c. We must now convert the LBNL data, measured in kilowatts (kW) of capacity, to panel area. As solar insolation data are recorded in kWh per square meter per day, we must take a stand on a conversion between panel area and power rating. Following the dimensional analysis procedure outlined in the National Renewable Energy Laboratory (NREL)[9] solar module, we calculate effective panel area as follows:

$$\underbrace{Panel\,Area_{ics}}_{m^2} = \underbrace{Rating_{ics}}_{kW} \times \frac{1m^2}{1kW} \times \frac{1}{\chi_{ics}}$$

With area in hand we can calculate the annual solar generation by each system in kilowatt-hours per year.

$$solar_{ics} = Panel\,Area_{ics} \times \chi_{ics} \times \bar{\phi}_{cs} \equiv Rating_{ics} \times \bar{\phi}_{cs}$$

For panels missing an efficiency listing (approximately 25 percent of observations), we use the median efficiency rating for all panels installed that year, in that state[10]. Observations that are almost certainly coded incorrectly (negative efficiency values) are also dropped at this point. Average household solar generation at the county level, $solar_{cs}$ is then taken as the total generation of solar energy from residential systems across all households, divided by the total number of households in county c, HH_{cs}.

$$solar_{cs} = \frac{1}{HH_{cs}} \sum_{i=1}^{HH_{cs}} solar_{ics}$$

Since consumers have no control over the weather, they have no control over $\bar{\phi}_{cs}$, so we

[8] After dropping observations due to incomplete observations from either missing geo-spatial indices or missing pricing variables as described in Section 3.4, we have enough data to compute empirical aggregates of residential solar panel electrical generation for 372 U.S. counties. The counties that remain after selection are the non-gray counties in Figure 3.

[9] https://pvwatts.nrel.gov/pvwatts.php

[10] Refining the replacement efficiency to state level or using a broader national median does little to affect the interpolation here.

ultimately care about the average panel rating, which is just:

$$Rating_{cs} = \frac{solar_{cs}}{\bar{\phi}_{cs}} \equiv \frac{1}{HH_{cs}} \sum_{i=1}^{HH_{cs}} PanelArea_{ics} \times \chi_{ics}$$

3.3 Costs and Incentives at the County Level

Both gross cost and incentive programs for residential SPV installation display substantial variation at the county level. North Carolina State University's DSIRE program indicates there have been 565 regulatory policies or financial incentives put in place since the year 2000 at the state and local level.[11] Regulatory actions that incentivize SPV uptake include easements for household installations, right-to-install laws, net metering, building requirements, and expediting permits. Financial incentives typically take the form of production-based incentives (subsidies that are determined based on households' generation level) or one-time transfers, such as rebates, grants, or tax credits.

Observations in the LBNL Tracking the Sun dataset record three fiscal incentives: 1) rebate or grant programs (R); 2) Performance-Based Incentives (PBI); 3) Feed-In Tariffs (FIT). We observe the magnitude and duration of these fiscal instruments for over 1 million residential SPV installations since 2000. Additional consideration is given to the Investment Tax Credit (ITC): households that installed panels between 2006 and 2020 may claim 30% of costs as a tax credit against federal income taxes.[12]

While PBI and FIT subsidies are very limited (only 0.0002% of our sample are observed to have received positive FIT and only 0.2% positive PBI), one-time installation rebates have been widely deployed to incentivize residential solar uptake. Figure 3 shows the geo-spatial distribution of average rebates $\widehat{R}_{cs}$ less local taxes $\widehat{T}_{cs}$ associated with panel purchase and installation by county from the LBNL data selected after cleaning.[13] Counties in Florida and Texas feature the greatest level of rebates. Given this, we would expect excess residential panel uptake conditional on production potential relative to what would otherwise be predicted, but that is not what we observe. In Section 3.5 we construct an index that characterizes the geo-spatial efficiency of residential solar installations across the U.S. Relative to counties in the west and northeast, we will show that despite large rebates, counties in Texas and Florida have an excess supply of un-tapped solar potential given conditionally lower rates of panel installation.

[11]See `https://programs.dsireusa.org/system/program`.

[12]See `https://www.energy.gov/eere/solar/downloads/residential-and-commercial-itc-factsheets`.

[13]The grey counties in Figure 3 do not contain enough installation observations to accurately estimate the local value of residential solar panels.

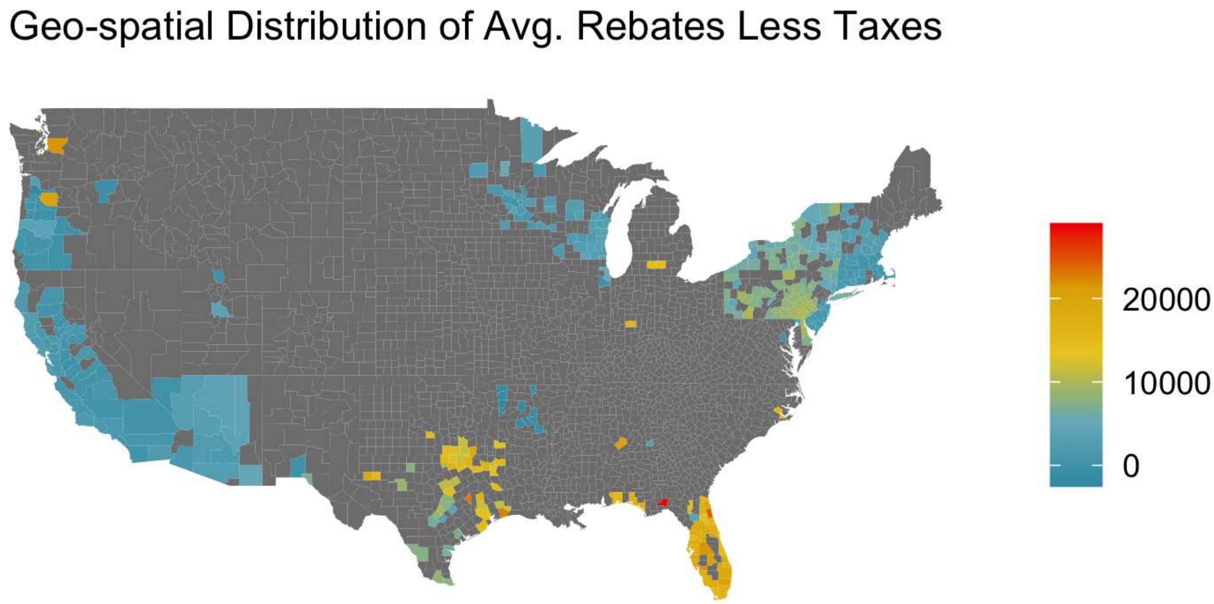


Figure 3: Here, we present average rebates less local sales taxes, $\widehat{R}_{cs} - \widehat{T}_{cs}$, by county in nominal dollars. Net rebates are highest in Texas and Florida, though King County Washington (home of Seattle) also has offered relatively high net rebates over the past couple of decades.

Data from LBNL are also robust on the cost side. All observations contain incentive structure data, installation prices and system size along with either reported or estimated sales taxes collected by local governmental authorities. This gives us a direct measure of the tax-inclusive cost of installation, allowing for a measure of total cost for each observation. Assuming a maximal uptake of ITC claims by households contributes an additional \$9.5 billion to subsidies over this period. Under the maximal ITC claims assumption, the combination of rebates and tax expenditures amount to greater than 64 percent of total residential SPV installation costs, tax inclusive, during the first decade of the 2000s. In the next section we will use the rebate, tax, and cost information featured in the LBNL data to estimate a price index that describes the local, county-level value of residential solar electrical consumption.

3.4 Constructing the Price of Solar Energy

To construct the implicit price of one kWh of energy derived from residential solar, we must account for local subsidization policies, as well as the rate of long-run depreciation for SPVs.[14] The LBNL dataset allows us to account for these factors under a couple of assumptions. Specifically, we calculate p_{cs}^{solar} at the county level using a levelized cost of energy (LCOE) approach, following a simplified version of Flowers et al. (2016). The LCOE is the ratio of the present value of lifetime net expenditures (costs minus subsidies) to the present value of electricity generation (the sum of lifetime discounted generation). The LBNL dataset provides data on installation costs, sales taxes, lump-sum rebates or grants, and levels and durations of FIT/PBI programs. The remaining parameters required to calculate LCOEs are assigned according to established values in the literature. We assume a 30-year lifespan for panels, a 1% annual depreciation rate (DR), an annual cost of maintenance equal to 1% of fixed costs (C_{ics}), and a discount rate of 3% (r), in line with Jordan and Kurtz (2013) and Flowers et al. (2016). Further, lacking data on whether or not individual households used ITC tax credits, we construct two separate solar price variables. In the first, which we use in our primary structural estimation and counterfactual exercises, we construct p_{cs}^{solar} under the assumption that ITC uptake is zero. In the second, we allow for maximal uptake of ITC. Estimates of our structural model under the two different price indices are similar, so we do not perform counterfactual analyses with the price indices where ITC uptake is assumed maximal.

Price construction induces further sample selection that, after aggregation, gets us to the final count of 372 counties containing usable aggregate solar demand observations. Many panel installation observations from the LBNL data are missing entries for either the installation price, whether a grant/rebate for the installation was received, or the amount of FITs and PBIs. The incidence of missing price variables varies greatly among states, and comprises about 25% of all observations. The states of Ohio, New Mexico, Montana, Kansas, and the District of Columbia feature all observations that are missing at least some of the required pricing variables, so we drop these states from our sample. To allow us to keep a larger share of the remaining observations missing price data, we assign average LCOEs at the county-year level to observations in that same county that are missing prices. Finally, we also drop counties for which p_{cs}^{solar} is implied to be negative. This can happen as a result of observing very few installations, all of which are highly subsidized, thus forcing $LCOE_{ics} < 0$.

In addition to fixed costs C_{ics}, let T_{ics} represent local sales taxes, and let R_{ics} be the one-time local rebate received upon installation. LCOEs must be calculated to account for the fact that FITs and PBIs expire after so many periods. The rates and dates of expiry depend on local policies as well as when the panel system was acquired by the consumer

[14]This price encodes the net present value of one unit of residential solar energy consumed, accounting for all possible costs associated with purchasing, installing, and maintaining the panels.

and, sometimes, whether or not the consumer filed the necessary paperwork to receive FIT and/or PBI credits. Let τ_{ics}^{FIT} be the number of periods after which FITs expire and τ_{ics}^{PBI} be the number of periods after which PBIs expire. Given a 30-year lifespan, the LCOE is:

$$LCOE_{ics} = \frac{C_{ics} + T_{ics} - R_{ics} + \sum_{t=0}^{29} \frac{0.01 \times C_{ics}}{(1+r)^t} - \sum_{t=0}^{\tau_{ics}^{FIT}} \frac{FIT_t (1-DR)^t solar_{ics}}{(1+r)^t} - \sum_{t=0}^{\tau_{ics}^{PBI}} \frac{PBI_t (1-DR)^t solar_{ics}}{(1+r)^t}}{\sum_{t=0}^{29} \frac{(1-DR)^t solar_{ics}}{(1+r)^t}} \quad (1)$$

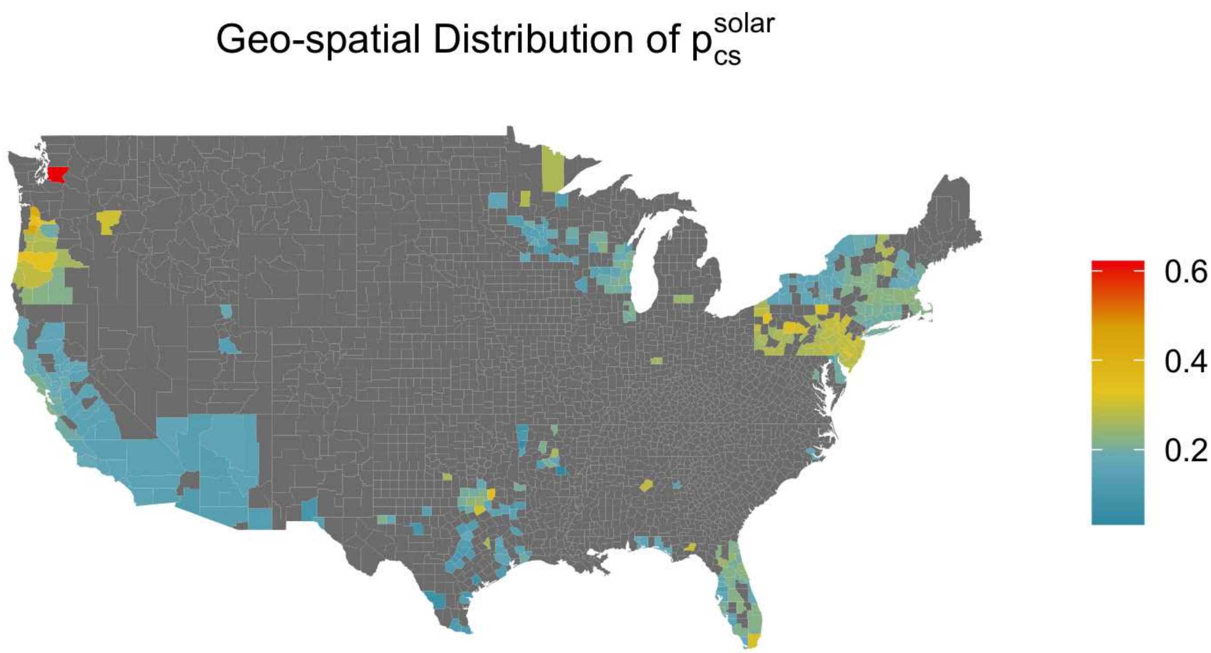


Figure 4: In this figure, we present a choropleth map of estimated local solar prices, assuming 0% ITC uptake. High $\overline{\phi}_{cs}$ and R_{ics} weigh negatively on $LCOE_{ics}$, thus driving down solar prices. Despite high rebates, King County Washington also is associated with the highest price of solar due to very low solar insolation.

Let $\widetilde{HH}_{cs}$ be the total number of household-level panel installation observations for county-state unit c, s. Note that $\widetilde{HH}_{cs} \leq HH_{cs}$ always, since we cannot observe more households with installed solar capacity than the total number of households that exist in the county. In fact, it is always true in our data that $\widetilde{HH}_{cs} < HH_{cs}$, strictly. Let $\widetilde{\tau}_{ics}$ be the year in which agent i who lives in county-state c, s purchased and installed solar panels. We calculate county-level average solar prices as the average of LCOE values across all installed units (converted to \$2018) and over time, weighted by system size S_{ics}:

$$p_{cs}^{solar} = \frac{\sum_{t=2000}^{2018} \sum_{i=1}^{\widetilde{HH}_{cs}} \mathbf{1}(\widetilde{\tau}_{ics} = t) S_{ics} LCOE_{ics} \frac{p_{2018}}{p_t}}{\sum_{t=2000}^{2018} \sum_{i=1}^{\widetilde{HH}_{cs}} S_{ics}} \tag{2}$$

Sums start at $t = 2000$ since that is the first year in our sample. p_t is the Consumer Price Index for All Urban Consumers (CPI-U) from the Bureau of Labor Statistics. In our baseline estimation, where we assume ITC uptake is zero, R_{ics} contains only local rebates. When we assume ITC uptake is maximal, we adjust R_{ics} to include 30% of costs — $0.3 \times C_{ics}$.

Figure 4 presents the geo-spatial distribution of solar prices. Given the LCOE calculations in (1), rebates weigh negatively on the price. Notice that despite issuing average net rebates of over \$10,000, King County, Washington (home to Seattle) is also associated with one of the highest effective solar prices, due to the combination of both high installation costs and low average daily insolation, $\overline{\phi}_{cs}$. Since $solar_{ics} = Rating_{ics} \times \overline{\phi}_{cs}$, low $\overline{\phi}_{cs}$ puts upward pressure on LCOE, and high $\overline{\phi}_{cs}$ puts downward pressure on LCOE. We see similar trends with respect to Oregon, which offers few rebates but also has low $\overline{\phi}_{cs}$ driving prices up. Pennsylvania, also, offers relatively high subsidies (around \$10,000 on average), and yet features relatively above-average solar prices due to relatively low daily insolation.

Examining variation between the geo-spatial net rebate and price distributions leads to a couple of questions. Rebates in cloudy places like Seattle seem poorly targeted from a national perspective, even though local policymakers may prefer them. Can we construct a measure of geo-spatial efficiency in order to understand where there may be surpluses or shortages of installed panels, relative to other localities? In the next section, we propose such an index. Second, given such potential aggregate, national inefficiencies, do local rebates contribute to a more efficient geo-spatial distribution of residential panels, relative to a no-rebate environment? We explore this second question with our structural counterfactuals in Section 6.

3.5 Geo-spatial Efficiency of Residential Coverage

As previously mentioned, the most efficient distribution of solar panels would see great concentrations of residential solar generation in the Southwest and West with sparser concentration in cloudier places like the Pacific Northwest and Northeast. However, variation in local subsidization of solar electricity and solar installation costs may incentivize consumers in places with relatively low average solar insolation to engage in relatively high rates of solar adoption. For this reason, we seek a measure of the degree to which the current distribution of residential panels across the U.S. deviates from the efficient distribution, which would follow the geo-spatial gradient in Figure 1, absent consideration for transmission frictions across a fractured electrical grid.

The share of average residential electrical consumption in county-state unit c, s that is attributed to residential solar is given by:

$$\mu_{cs} = \frac{Rating_{cs}\overline{\phi}_{cs}}{e_{cs}} \equiv \frac{solar_{cs}}{e_{cs}}$$

where e_{cs} is per-household total electrical consumption and $\mu_{cs} \in [0, 1)$, allowing for no residential solar consumption but requiring that some fraction of electrical consumption always comes from other sources. Let $\mathcal{C}$ be the number of county-state pairs in our sample. We can compare the residential solar market share of electrical consumption across these counties using the empirical distribution function:

$$\mathbb{P}_{\mu}(m) = \frac{1}{\mathcal{C}} \sum_{s} \sum_{c} \mathbf{1}_{\mu_{cs} \leq m}$$

We can also rank counties in terms of their average daily solar insolation:

$$\mathbb{P}_{\phi}(\varphi) = \frac{1}{\mathcal{C}} \sum_{s} \sum_{c} \mathbf{1}_{\overline{\phi}_{cs} \leq \varphi}$$

Using both of these statistics, we construct a unit-less index on the log-scale that characterizes the degree to which the geo-spatial distribution of residential solar panels deviates from the efficient distribution:

$$\lambda_{cs} = \ln\left(\frac{\mathbb{P}_{\mu}(\mu_{cs} = m)}{\mathbb{P}_{\phi}(\overline{\phi}_{cs} = \varphi)}\right) \tag{3}$$

If $\lambda_{cs} > 0$ then county-state c, s has excess installed residential solar capacity relative to the other counties in our sample. If $\lambda_{cs} < 0$ then county-state c, s has an excess supply of un-tapped solar potential relative to the other counties in our sample.

Note that the moments of the distribution of λ_{cs} provide rich summary statistics characterizing how the geo-spatial distribution deviates from the efficient distribution. As $\mathbb{E}(\lambda_{cs}) \to 0$ and $\text{Var}(\lambda_{cs}) \to 0$ simultaneously, the distribution of λ_{cs} converges to the efficient distribution. If $\mathbb{E}(\lambda_{cs}) = 0$ and $\text{Var}(\lambda_{cs}) > 0$ then the distribution is inefficient yet unbiased. That is, geo-spatial inefficiency is equally attributable to excess supply in cloudier places and relative shortages in sunnier places. If $\mathbb{E}(\lambda_{cs}) \neq 0$ but finite,

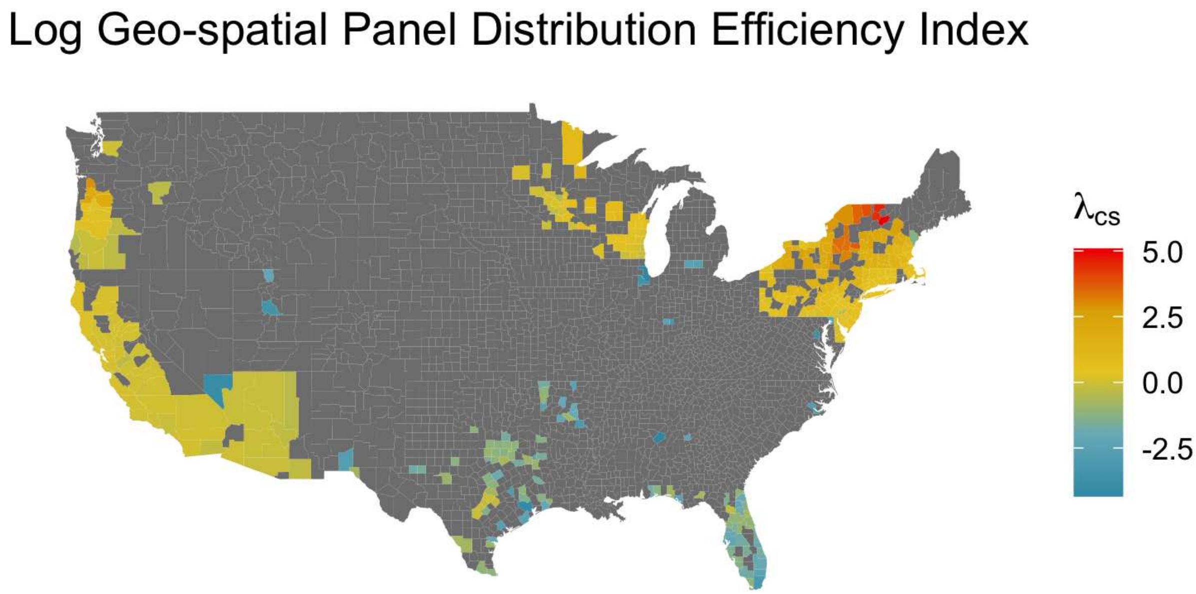


Figure 5: The geo-spatial distribution of the log efficiency index, λ_{cs}, shows that relative excess residential-solar supply is concentrated in the Northeast, while relative residential-solar shortages are concentrated primarily in Texas and Florida. The mean efficient county in our sample is Tuolumne County near Yosemite National Park in Central California.

then necessarily $\mathrm{Var}(\lambda_{cs}) > 0$, and the distribution is both inefficient and biased. If $\mathbb{E}(\lambda_{cs}) < 0$ then, generally speaking, residential solar capacity is inefficient due more to under-installation in sunnier states relative to the efficient distribution. If $\mathbb{E}(\lambda_{cs}) > 0$ then, generally speaking, residential solar capacity is inefficient due more to excess solar capacity in cloudier states relative to the efficient distribution.

After the sample selection procedures discussed in the previous sections, we can compute the geo-spatial distribution of λ_{cs} for the counties that survive our selection procedure. This geo-spatial distribution is presented in Figure 5, where redder counties feature a relative over supply of panels, while bluer counties feature relative under supply. Notice that over supply appears concentrated in the Northeast and Pacific Northwest, while counties in Texas are associated with a notable under supply considering the above average solar insolation levels these counties experience. With respect to the distribution of λ_{cs}, we find that the mean is not significantly different from zero, though $\mathrm{Var}(\lambda_{cs}) = 2.164 > 0$, suggesting an unbiased but inefficient distribution of solar panels.

4 Model of Household Demand

The model we propose operates at the U.S. county level. Index county units by $c \in \mathbb{C}$. Each county belongs to a state $s \in \mathbb{S}$. A consumer-unit i in county c and state s has basic consumption preferences over housing h_{ics} (in units of square feet), total household electrical consumption e_{ics} which includes solar (in units of kWh), an additional amenity preference for consumption characterized by the panel rating variable $Rating_{ics}$ (in units of kW, not kWh), and an outside commodity $\overline{c}_{ics}$ which captures all other consumption. All choices are assumed to be strictly positive and continuous, and all consumers are assumed to be price takers, meaning that their atomic decisions have negligible influence on general-equilibrium outcomes. In our preference formulation consumers have control over the overall kilowatt-rating of the panels they purchase but not their total output, which also depends on local insolation. The household decision environment is a static one: we do not consider savings or investment decisions but rather account for the long-run effects of solar-panel investment by constructing prices accordingly, as previously described. The goal of our modeling exercise is to find reasonable estimates of structural demand parameters in order to explore how both local and national subsidization policies may contribute to the inefficient geographic distribution of solar capacity.

We do not distinguish between electricity consumed from solar versus other sources. Instead, consumers prefer to consume solar electricity only because they have preferences for the solar amenity. That is, they like solar perhaps because they are environmentally conscious or enjoy the feeling of self-reliance derived from consuming electricity produced with equipment they own. However, a consumer, when in the act of consuming electricity to power appliances or lights, for example, does not care about the source of electricity. In this sense we assume that the solar amenity preference and preferences for consumption are somehow separable in utility.

The preference structure we impose is of the nested constant elasticity of substitution (CES) variety. Our nesting structure features a three-tiered hierarchy. To allow for the fact that housing consumption and electrical consumption may be complementary, we allow them to be contained within their own nest. The next layer accounts for complementarities between the composite of housing and electrical consumption and the solar amenity. Finally, the outer-most layer accounts for the relationship between housing/electrical consumption plus the solar-amenity composite and all other consumption:

$$u_{cs}(h_{ics}, e_{ics}, Rating_{ics}, \overline{c}_{ics}) =$$
$$\left[\left(\left(\gamma h_{ics}^{\alpha} + (1-\gamma)e_{ics}^{\alpha}\right)^{\frac{\rho}{\alpha}} + \eta_{cs}(Rating_{ics}\overline{\Phi}_{cs})^{\rho}\right)^{\frac{\kappa}{\rho}} + \overline{c}_{ics}^{\kappa}\right]^{\frac{1}{\kappa}}, \tag{4}$$
$$\alpha, \rho, \kappa \neq 0 \quad \text{and} \quad \alpha, \rho, \kappa < 1$$

α governs the degree to which housing and electricity are complementary. ρ governs the substitutability between the housing/electricity composite good and the solar amenity. κ governs the rate at which consumers substitute between all other consumption and the housing/electricity plus solar-amenity composite. γ governs the relative importance consumers place on housing versus electrical consumption. η_{cs} characterizes the degree to which local demographic, political, cultural, and climatic factors impact solar consumption preferences. Note that all consumers in the same county-state unit face the same amenity-preference weight η_{cs}. We assume that consumers value electricity independently of where it comes from but care about solar-panel-generated electricity only as an amenity. That is, they do not separately choose non-solar and solar electrical consumption. This assumption reflects the difficulty in decomposing electrical consumption from solar versus non-solar due to grid-based electricity comprising flows from both sources.

All households in the same county-state unit face the same market prices. Let p_{cs}^{h} be the average price per square foot of owner-occupied housing, as reported in the 2014-2018 American Community Survey (ACS) 5-year estimates. Let p_{cs}^{e} be the market price per kilowatt hour of electricity in county c and state s. Let p_{cs}^{solar} be the net-present-value of solar electricity consumption, accounting for wear-and-tear and depreciation of panel efficiency.[15] $p_{cs}^{\overline{c}}$ is a price index describing the value of all other consumption. Let y_{ics} encode total income net of savings for household i in county c and state s. The household's budget must satisfy

$$\begin{aligned} p_{cs}^{h}h_{ics} + p_{cs}^{e}e_{ics} + p_{cs}^{solar}Rating_{ics}\overline{\Phi}_{cs} + p_{cs}^{\overline{c}}\overline{c}_{ics} \\ \leq y_{ics} + \max\{p_{cs}^{e}(Rating_{ics}\overline{\Phi}_{cs} - e_{ics}), 0\} \end{aligned} \tag{5}$$

Within a county-state unit, the exogenous source of household heterogeneity is thus limited to y_{ics}. Note that in most localities, excess electricity generated from residential solar panels is sent back to the grid, and the household is offered a credit at par with the market-value of excess production, hence the final term on the right-hand-side.

[15] This is the solar price constructed in Section 3.4.

Since all household heterogeneity within a particular county-state unit is governed by exogenous variation in income, we can exploit the homotheticity of u_{cs} to focus solely on the problem of a representative consumer, given u_{cs} is associated with an indirect utility function that satisfies the polar form of Gorman (1959). Further, since we observe that $\sum_i (Rating_{ics}\overline{\Phi}_{cs} - e_{ics}) < 0$ in all county-state combinations in our sample, it follows that no county produces more residential solar electricity than the total amount of electricity it consumes. Thus, we can safely ignore the argument $\max\{p^e_{cs}(Rating_{ics}\overline{\Phi}_{cs} - e_{ics}), 0\}$ on the right-hand side of the representative consumer's budget constraint when characterizing aggregate demand properties in the next section. From here on, we drop i subscripts. Moving forward, we will let choice variables that are missing their i subscripts correspond to aggregate demands for county c in state s.

4.1 Equilibrium Aggregate Demand

The representative consumer in county c and state s solves:

$$\begin{aligned} \max_{h_{cs}, e_{cs}, Rating_{cs}, \bar{c}_{cs}} \quad & u_{cs}(h_{cs}, e_{cs}, Rating_{cs}, \bar{c}_{cs}) \\ \text{subject to} \quad & p^h_{cs}h_{cs} + p^e_{cs}e_{cs} + p^{solar}_{cs} Rating_{cs}\overline{\Phi}_{cs} + p^{\bar{c}}_{cs}\bar{c}_{cs} \leq y_{cs} \end{aligned}$$

In equilibrium, the marginal rate of substitution between housing and total electrical consumption is:

$$\frac{\gamma}{1-\gamma}\left(\frac{h_{cs}}{e_{cs}}\right)^{\alpha-1} = \frac{p^h_{cs}}{p^e_{cs}} \tag{6}$$

The marginal rate of substitution between consuming the solar amenity and total electricity is:

$$\frac{\eta_{cs}\overline{\Phi}^{\rho}_{cs} Rating^{\rho-1}_{cs}}{\left(\gamma h^{\alpha}_{cs} + (1-\gamma)e^{\alpha}_{cs}\right)^{\frac{\rho-\alpha}{\alpha}}(1-\gamma)e^{\alpha-1}_{cs}} = \frac{p^{solar}_{cs}\overline{\Phi}_{cs}}{p^e_{cs}} \tag{7}$$

Given the separability between the nest which includes h_{cs}, e_{cs}, and $Rating_{cs}$ and $\bar{c}_{cs}$, we can fully characterize the demand for housing, electricity, and solar panels without a need to specify equilibrium first-order conditions for $\bar{c}_{cs}$. Indeed, separability between $Rating_{cs}$ and the housing/electricity nest allows us to substitute out h_{cs} from (7) using (6) in order to arrive at an expression that roughly characterizes the fraction of electrical consumption devoted to solar:

$$\left(\frac{Rating_{cs}}{e_{cs}}\right)^{\rho-1}\frac{\eta_{cs}\overline{\phi}_{cs}^{\rho}}{1-\gamma}\left(\gamma\left[\frac{p_{cs}^{h}(1-\gamma)}{p_{cs}^{e}\gamma}\right]^{\frac{\alpha}{\alpha-1}}+1-\gamma\right)^{\frac{\alpha-\rho}{\alpha}}=\frac{p_{cs}^{solar}\overline{\phi}_{cs}}{p_{cs}^{e}} \tag{8}$$

(8) can be used in order to arrive at consistent estimates of the structural parameters, given data for prices, demand, and average daily insolation by county. (8) is thus our primary structural estimating equation around which we will form a likelihood function and arrive at estimates of α, ρ, γ and a residual expression for the amenity weight η_{cs}. Given κ does not factor into our primary estimating equation, which characterizes the marginal rate of substitution between housing/electricity and the solar amenity, we need not estimate this parameter, so it remains free. In the following section, we provide the details of our structural estimation approach, while also addressing the orthogonality assumptions required for consistent recovery of the parameters.

5 Structural Estimation Procedure

Our estimation procedure operates on county-level data for 372 counties, distributed mostly in the western U.S., northern Midwest, Texas, Florida, and the northeastern U.S. We select a sample of counties for which we have sufficient data on residential solar panel installations, including cost of installation and any rebates, feed-in tariffs, or performance-based incentives received by the household, where county selection is as described in Section 3.

The primary equation we target is a logged version of (8), controlling for local demographic, economic, and environmental factors that may be correlated with the amenity weight, η_{cs}. Our approach is to form a likelihood function around (8) while placing prior distributions (flat, where possible) on the structural parameters, ρ, α, and γ, as well as the likelihood variance σ^2, in order to estimate the posterior distribution of model parameters given data. We thus take a Bayesian approach to estimation and use Monte Carlo integration techniques to estimate the posterior distribution. Specifically, we estimate the posterior distribution using a Hamiltonian Monte Carlo (HMC) method as described in Neal (2011). We run four different versions of the estimation procedure, separately restricting ρ and α to either be in $(0, 1)$ or $(-\infty, 0)$. We then employ model selection techniques to compare the log-posterior predictive density functions of the separately-estimated models, where sampling from the predictive distribution is undertaken using the leave-one-out methods described in Vehtari, Gelman, and Gabry (2017).

The estimation results presented in the main text involve estimates of the solar price index where, lacking data, we assume that uptake of federal ITC is zero. In Appendix A we present estimation results under the alternative assumption that ITC uptake is the maximum 30% for all observed installations. There is no significant discernible difference between the posterior distribution results under either price-construction assumption, so we dispense with performing counterfactual analyses on the model where p_{cs}^{solar} is constructed allowing for 30% ITC uptake.

Our parameter estimates and model selection procedures suggest that $\rho \in (0, 1)$, so that the solar amenity and the housing/electricity composite are substitutes. We also prefer the model where $\alpha < 0$, which we believe to be consistent with theory: we expect housing and electricity to be complementary since larger houses are to be associated with more appliances, more lights, and more space to heat and cool, and thus greater electrical consumption. Controlling for local demographic, political, cultural, and climatic factors, we find no significant relationship between solar demand and either population density, median household income, or rates of home ownership. We find that counties with higher shares of Republican voters (GOP voting share) are somewhat surprisingly associated with greater preferences for the solar amenity. Meanwhile, the number of cooling degree days is the most significant local factor weighing down on solar amenity preferences. We discuss all of these findings and present robustness tests to check for regressor endogeneity in the sections that follow.

5.1 Likelihood and Prior Distributional Assumptions

Taking logs of (8) and isolating the log of the amenity weight as a residual, $\ln \eta_{cs}$, we get:

$$\begin{aligned} \ln\left(\frac{p_{cs}^{solar}}{p_{cs}^{e}}\right) + (1-\rho)\ln\left(\frac{Rating_{cs}}{e_{cs}}\right) + (1-\rho)\ln\overline{\Phi}_{cs} + \ln(1-\gamma) \\ + \left(\frac{\rho-\alpha}{\alpha}\right)\ln\left(\gamma\left[\frac{p_{cs}^{h}(1-\gamma)}{p_{cs}^{e}\gamma}\right]^{\frac{\alpha}{\alpha-1}} + 1-\gamma\right) = \ln\eta_{cs} \end{aligned} \tag{9}$$

We expect the amenity weight η_{cs} to be correlated with local demographic, economic, and environmental factors, such as population density, median household income, political preferences, and the number of hot versus cold days during the year. We thus write

$\ln \eta_{cs}$ as a linear function of the following set of controls:

$$\begin{aligned}\ln \eta_{cs} = {} & \beta_1 \ln pop\_density_{cs} + \beta_2 \ln GOP\_voting\_share_{cs} \\ & + \beta_3 \ln median\_income_{cs} + \beta_4 \ln heating\_days_{cs} \\ & + \beta_5 \ln cooling\_days_{cs} + \beta_6 \ln homeownership\_rate_{cs} + \epsilon_{cs}\end{aligned}$$

Population density is from the 2011-2015 ACS 5-year estimates. The GOP voting share variable measures the share of the vote the Republican candidate, Donald Trump, received in the 2016 presidential election. Median income is the median income of a home-owning household, not a person, from the 2014-2018 ACS 5-year estimates. Heating and cooling degree days are measures which quantify a given locale's demand for energy to heat and cool buildings, depending on the region's daily outdoor air temperature.[16] These variables are found in the DeepSolar data along with $\overline{\phi}_{cs}$. The homeownership rate describes the fraction of all households that are owner-occupied, as found in the 2011-2015 ACS 5-year estimates. ϵ_{cs} is the idiosyncratic residual that is assumed *iid* across county-state observations. We form the likelihood function around this error term, which is assumed normally distributed, and the variance of which is given a conjugate inverse gamma prior:

$$\begin{aligned}&\epsilon_{cs} \sim_{iid} \mathcal{N}(0, \sigma^2) \\ &\sigma^2 \sim \text{InvGamma}(5, 5)\end{aligned}$$

The structural preference parameters ρ and α are given priors that depend on whether we ex-ante (prior to running the integration procedure) restrict ρ and/or α to be in the unit interval or negative:

$$\begin{aligned}&\rho, \alpha \sim \mathcal{U}[0,1], \quad \text{if either } \ \rho \ \text{ and/or } \ \alpha \in (0,1) \\ &-\rho, -\alpha \sim \mathcal{LN}\left(-\frac{1}{2}, 1\right), \quad \text{if either } \ \rho \ \text{ and/or } \ \alpha < 0\end{aligned}$$

$\mathcal{U}$ is the uniform distribution and $\mathcal{LN}$ is the log-normal distribution. Note that we run four separate estimations accounting for the fact that both ρ and α may be positive, one is negative and the other is positive, or both are negative.

[16]To compute heating degree days, for example, take a region's average temperature on any given day and subtract it from a pre-set base temperature, which is effectively the constant building temperature that is desired. This constant temperature is attributed to local building standards and is typically taken to be 18.3°C (65°F) in the United States. If this value is positive, then the value gives the number of heating degree days. If it is negative, the absolute value yields the number of cooling degree days.

The contribution of housing consumption to the housing/electricity preference nest is given by γ. Having found in simulations that this parameter often piles up at its upper bound of 1, we impose an informative beta prior on γ:

$$\gamma \sim \text{Beta}(1, \omega)$$

ω is a hyper-parameter which we estimate by taking logs of the infra-marginal rate of substitution between solar and electrical consumption in (6) and running an OLS regression to get a prior estimate for γ which we call $\widehat{\gamma}$. We allow this prior estimate to correspond to the mean of the beta distribution, which gives us $\widehat{\omega} = 1/\widehat{\gamma} - 1$. This value is $\widehat{\omega} = 0.289$.

Finally, the contribution of control variables to the amenity weight is governed by the vector $\boldsymbol{\beta}$ with components β_j. We give these parameters flat, improper priors, allowing them to take any real value with equal quasi-probability:

$$\beta_j \sim \mathcal{U}(-\infty, \infty), \quad \forall j \in \{1, 2, 3, 4, 5, 6\}$$

While improper priors can sometimes create computational issues when sampling, convergence criteria for our integration schemes are all satisfied and no such issues arise.

5.2 Posterior Distribution Estimates and Model Selection

Table 1: Model Selection Assessments

Model	$elppd$ Difference[a]	S.E. Difference[b]
$\rho, \alpha \in (0, 1)$	—	—
$\rho \in (0, 1), \alpha < 0$	-11.2	3.3
$\rho < 0, \alpha \in (0, 1)$	-797.5	21.8
$\rho, \alpha < 0$	-799.0	21.7

[a] This is the difference in the expected log-posterior predictive density relative to the model with the highest $elppd$ value — $\rho, \alpha \in (0, 1)$. As long as the absolute value of the $elppd$ difference is several times the standard error of the differences, we can be confident that the baseline model provides a better fit than its alternatives (Vehtari, Gelman, and Gabry 2017).

[b] This is the standard error of the posterior differences across HMC sample draws.

The structural models, conditional on prior assumptions regarding how the solar amenity, housing, and electricity are related (substitutes versus complements), are estimated using an HMC integration routine parallelized over four independent chains. Each chain generates 8000 samples from the posterior distribution of parameters conditional upon prior assumptions and data. We allow the first 4000 samples to be considered warm-up/burn-in samples and drop them from our posterior analyses, leaving us with $4000 \times 4 = 16000$ sample parameter vectors for each model. The operations required to perform the HMC take approximately 20 minutes on a standard multi-threaded, multi-core laptop. All four models converge according to the $\widehat{R}$ convergence criterion for HMC estimation procedures, as outlined in Vehtari et al. (2019).[17]

Tables 2 and 3 present the posterior distribution estimates of the model's structural parameters, under the respective assumptions that $\rho \in (0,1)$ and $\rho < 0$. The top half of each table contains estimates under the assumption $\alpha \in (0,1)$, while the bottom half of each table contains estimates assuming $\alpha < 0$. We perform model selection by assessing the expected log posterior predictive density ($elppd$) using the leave-one-out sampling method as described in Vehtari, Gelman, and Gabry (2017). $elppd$ is largest for the model with $\rho, \alpha \in (0,1)$, though for theoretical reasons we prefer the model where $\rho \in (0,1)$ and $\alpha < 0$, allowing housing size and electrical consumption to be complements. As can be seen in Table 1 the model with $\rho, \alpha \in (0,1)$ is only weakly preferred, in terms of $elppd$ to that with $\rho \in (0,1)$ and $\alpha < 0$, relative to the two models where $\rho < 0$. Since both models where $\rho \in (0,1)$ are strongly preferred, in terms of $elppd$ to those where $\rho < 0$, solar demand and demand for housing/electrical consumption are substitutes. Further evidence that the $\rho \in (0,1)$ models provide better fitness can be seen by comparing the likelihood variances σ^2 in Table 2 to those in Table 3. The likelihood spreads are substantially larger in the complements models with $\rho < 0$. The log-posterior densities $V(\mathcal{P})$ are also larger in models with $\rho \in (0,1)$, so that the posterior is tighter and more concentrated than the estimated distributions in models with $\rho < 0$. For all of these reasons, we focus our analyses on the estimates in Table 2.

[17]The criterion states that if the $\widehat{R}$ statistics for all model parameters as well as the same statistic associated with estimates of the log posterior density $V(\mathcal{P})$ are approximately unity, then the numerical integration routine has converged to the true posterior distribution. The $\widehat{R}$ statistic used to assess HMC integration convergence should not be confused with (and is unrelated to) the R^2 statistic used by frequentists in ANOVA tables.

Table 2: Posterior Distribution Estimates, $\rho \in (0,1)$

$\alpha \in (0,1)$

Parameter	$\widehat{R}$	Mean	S.D.	2.5%	25%	50%	75%	97.5%
ρ	1.000	0.951	0.007	0.937	0.946	0.951	0.955	0.965
α	1.001	0.748	0.106	0.473	0.701	0.772	0.822	0.883
γ	1.000	0.967	0.027	0.900	0.963	0.974	0.980	0.987
σ^2	1.000	0.104	0.008	0.090	0.099	0.104	0.109	0.121
β_1	1.000	0.010	0.024	-0.037	-0.006	0.010	0.026	0.058
β_2	1.000	0.181	0.063	0.056	0.138	0.181	0.224	0.304
β_3	1.000	-0.034	0.073	-0.176	-0.082	-0.034	0.014	0.110
β_4	1.000	-0.104	0.038	-0.177	-0.129	-0.104	-0.078	-0.029
β_5	1.000	-0.419	0.074	-0.565	-0.469	-0.419	-0.370	-0.271
β_6	1.000	-0.171	0.146	-0.455	-0.269	-0.173	-0.073	0.118
$V(\mathcal{P})^a$	1.001	235.024	2.363	229.397	233.735	235.397	236.736	238.531

$\alpha < 0$

Parameter	$\widehat{R}$	Mean	S.D.	2.5%	25%	50%	75%	97.5%
ρ	1.000	0.946	0.007	0.933	0.941	0.946	0.951	0.960
α	1.000	-0.171	0.116	-0.463	-0.219	-0.144	-0.091	-0.035
γ	1.000	0.364	0.124	0.111	0.280	0.370	0.451	0.595
σ^2	1.000	0.110	0.008	0.095	0.104	0.109	0.115	0.126
β_1	1.000	0.000	0.024	-0.048	-0.017	-0.000	0.016	0.047
β_2	1.000	0.189	0.066	0.060	0.145	0.189	0.234	0.318
β_3	1.000	0.058	0.084	-0.106	0.003	0.059	0.114	0.223
β_4	1.000	-0.085	0.038	-0.159	-0.111	-0.085	-0.060	-0.009
β_5	1.000	-0.361	0.075	-0.509	-0.412	-0.363	-0.311	-0.214
β_6	1.000	-0.157	0.149	-0.450	-0.256	-0.160	-0.056	0.135
$V(\mathcal{P})^a$	1.000	226.328	2.283	221.039	225.016	226.663	228.009	229.769

[a] $V(\mathcal{P})$ is the log posterior density of parameters, $\mathcal{P}$.

Table 3: Posterior Distribution Estimates, $\rho < 0$

$\alpha \in (0,1)$								
Parameter	$\widehat{R}$	Mean	S.D.	2.5%	25%	50%	75%	97.5%
ρ	1.000	-0.011	0.005	-0.022	-0.013	-0.010	-0.007	-0.004
α	1.000	0.570	0.264	0.049	0.361	0.618	0.799	0.952
γ	1.000	0.784	0.257	0.095	0.688	0.915	0.963	0.994
σ^2	1.001	5.872	0.442	5.071	5.561	5.849	6.159	6.805
β_1	1.000	-0.186	0.174	-0.528	-0.304	-0.185	-0.065	0.153
β_2	1.000	-1.361	0.469	-2.280	-1.676	-1.356	-1.049	-0.435
β_3	1.000	0.722	0.518	-0.296	0.371	0.718	1.070	1.746
β_4	1.000	-0.324	0.278	-0.873	-0.512	-0.319	-0.137	0.221
β_5	1.000	-2.196	0.529	-3.234	-2.552	-2.192	-1.837	-1.173
β_6	1.000	3.202	1.121	1.009	2.459	3.209	3.957	5.405
$V(\mathcal{P})^a$	1.000	-540.145	2.632	-545.942	-541.733	-539.940	-538.316	-535.571
$\alpha < 0$								
Parameter	$\widehat{R}$	Mean	S.D.	2.5%	25%	50%	75%	97.5%
ρ	1.000	-0.011	0.005	-0.022	-0.013	-0.010	-0.007	-0.004
α	1.000	-0.483	0.558	-1.804	-0.587	-0.332	-0.185	-0.060
γ	1.000	0.645	0.316	0.033	0.385	0.727	0.942	1.000
σ^2	1.000	5.938	0.436	5.142	5.638	5.917	6.213	6.866
β_1	1.000	-0.237	0.175	-0.582	-0.355	-0.237	-0.118	0.099
β_2	1.000	-1.324	0.477	-2.264	-1.647	-1.323	-1.002	-0.395
β_3	1.000	0.461	0.526	-0.579	0.112	0.466	0.813	1.475
β_4	1.000	-0.198	0.261	-0.709	-0.371	-0.200	-0.022	0.317
β_5	1.000	-2.086	0.521	-3.117	-2.433	-2.085	-1.738	-1.062
β_6	1.000	3.125	1.078	1.030	2.399	3.116	3.857	5.248
$V(\mathcal{P})^a$	1.000	-541.208	2.319	-546.558	-542.526	-540.882	-539.530	-537.712

[a] $V(\mathcal{P})$ is the log posterior density of parameters, $\mathcal{P}$.

To better understand why we prefer the model with $\rho \in (0,1)$ and $\alpha < 0$, despite the model selection mechanism preferring $\rho, \alpha \in (0,1)$, let us compare estimates between the top and bottom half of Table 2. Note first that changing the prior distributional assumption on α does not substantially impact the posterior distribution of ρ. Rather, changing the prior on α appears to most significantly impact inference regarding the housing weight γ and the contribution of the control variables to amenity preferences $\boldsymbol{\beta}$. One reason we prefer the model where $\alpha < 0$ is that in that model $\gamma << 1$, compared to the model where $\alpha \in (0,1)$ and γ is close to 1. γ near 1 suggests that consumers do not care much about consuming electricity but really care about living in a big house. Further, $\alpha \in (0,1)$ suggests that house size and electrical consumption are themselves substitutes, so that, on the margin, consumers may increase their housing size but decrease their electrical consumption in response to an increase in electrical prices. This seems counter-intuitive: we would not normally expect consumers to demand larger houses in the middle of, for example, an energy crisis. Larger houses have more lights, more room for electronic devices and appliances, and require more energy for heating and cooling. Thus, while $elppd$ is weakly better for $\rho, \alpha \in (0,1)$, common sense suggests $\alpha < 0$ and so we prefer the specification in the bottom half of Table 2.

We will now focus on the values of $\boldsymbol{\beta}$ in the bottom half of Table 2. Population density (β_1) and median household income (β_3) appear to have no significant effect on local solar preferences. Meanwhile, GOP voting share (β_2) weighs positively on local preferences. Heating degree days (β_4), cooling degree days (β_5), and the homeownership share (β_6) all weigh negatively on local preferences, with the largest most significant effect coming from cooling degree days. At first glance we would expect the homeownership share to be positively correlated with solar amenity preferences, since county-aggregate residential solar market penetration relies on a high level of homeowners in the county. However, homeownership shares are highest in rural areas, and residential solar appears to have its greatest local market penetration on the urban coasts, especially the northeast and west coast. Further, we are somewhat surprised that Republican Party (GOP) voting shares are positively correlated with solar amenity preferences, given conservative hostility to climate change mitigation policies. We note, however, that the effect, while positive, is not strong: on average a 1% increase in the share of GOP voters in a county leads to an expected increase of 0.19% to η_{cs}.

Focussing still on the bottom half of Table 2 we turn our attention to coefficients describing the effects of heating degree days (β_4) and cooling degree days (β_5) on amenity preferences. Colder climates are associated with more heating degree days. Notice that the marginal posterior distribution of β_4 is centered closer to zero than that of β_5, demonstrating that the negative effect of more cooling degree days is stronger than that of higher heating degree days. What might be driving these relationships and differentials? First, as we see in Figure 5, excess residential solar supply relative to the efficient geo-distribution is mostly concentrated in colder parts of the country — the Pacific Northwest, New York state, and New England. Naturally, these areas have higher annual heating degree days. Since energy for heating buildings is often supplied by natural gas or heating oil, electricity demand in these regions will not be as adversely subjected to wild swings in grid-supply due to an over-reliance on solar energy. Second, we observe a relative under-supply of solar in hot and dry climates of the Southwest and Texas, as well as some more humid areas on the Gulf Coast of Florida. These areas are associated with high numbers of cooling degree days per year. Since air conditioning is almost always powered by electricity, in such places that require lots of electricity if grid supply relies on a large distribution of residential solar panels, then the grid could be stressed by over-demand on cloudier days.

5.3 Amenity Preferences and Idiosyncratic Errors

In this section we examine the structure of the error and amenity weight distributions. First, we examine the geo-spatial distributions of the posterior mean of the idiosyncratic errors $\widehat{\epsilon}_{cs}$ and the posterior mean of the log amenity weight $\widehat{\ln(\eta_{cs})}$ and compare them to the log efficiency index λ_{cs}. Second, we consider whether the error distribution itself may be systematically correlated with any of the regression variables, suggesting possible endogeneity bias. Third and finally, we test whether the error structure is correlated with local average subsidization policies to understand if price endogeneity due to systematic variation in subsidies could be affecting estimation results.

5.3.1 Correlation Between Errors, Amenity Preferences, and λ_{cs}

Are the geo-spatial distributions of the idiosyncratic errors and amenity weights correlated with the log-efficiency index? Figures 6, 7, 8a, and 8b help visually characterize these relationships for the model where $\alpha < 0$ and $\rho \in (0, 1)$. The geo-spatial distributions are presented in Figures 6 and 7, while scatterplots showing the relation between λ_{cs} and both $\widehat{\epsilon}_{cs}$ and $\widehat{\ln \eta_{cs}}$ are presented in Figures 8a and 8b. We find very little correlation (Pearson's coefficient is 0.10) between $\widehat{\epsilon}_{cs}$ and λ_{cs} and only slight linear correlation between $\widehat{\ln \eta_{cs}}$ and λ_{cs} (Pearson's coefficient is 0.27). These relationships can be seen in Figure 8. Meanwhile, Figure 7 shows that cooler, cloudier locales in the Pacific Northwest appear to have the largest solar-amenity preference weight. This is not surprising given the excess uptake of solar panels in that region with relatively little local subsidization, suggesting that the local demand for solar is not simply driven by low prices.

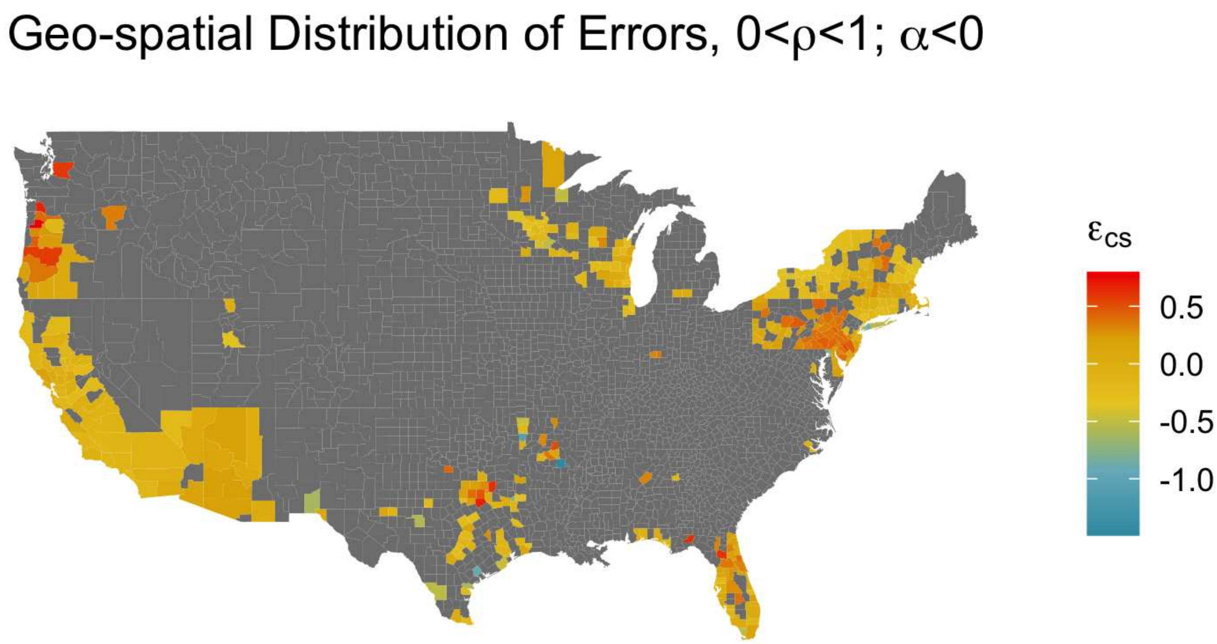


Figure 6: In this figure we show the geo-spatial distribution of the posterior means of structural idiosyncratic errors, $\widehat{\epsilon}_{cs}$, for all counties in our sample.

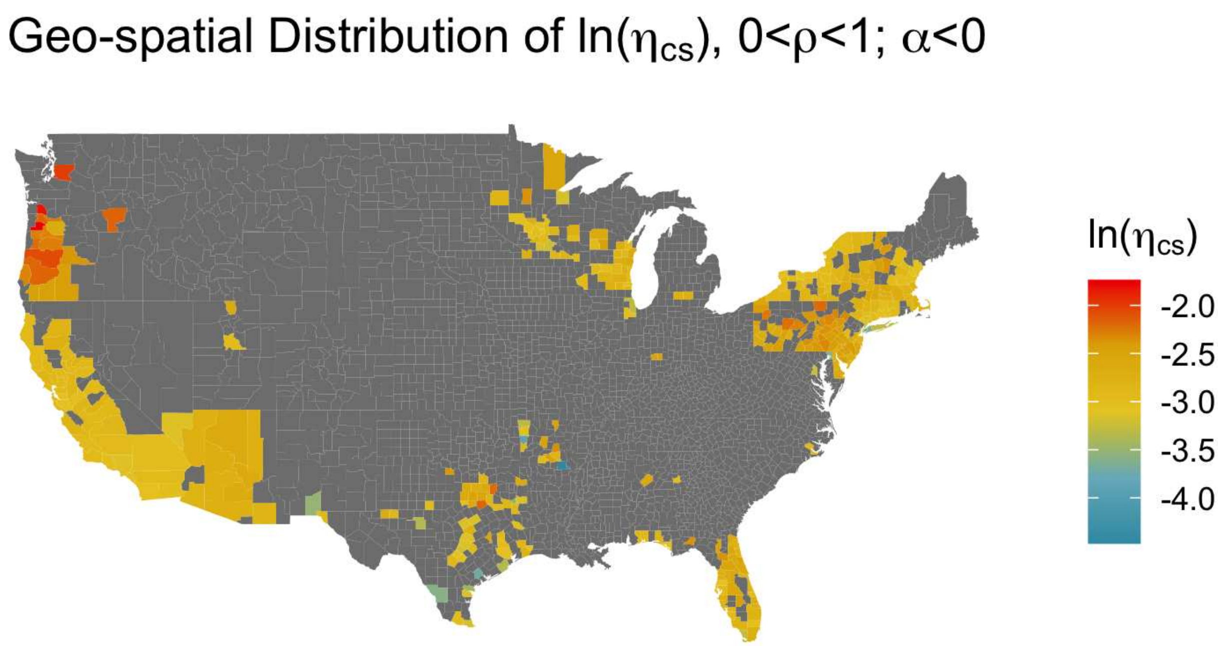
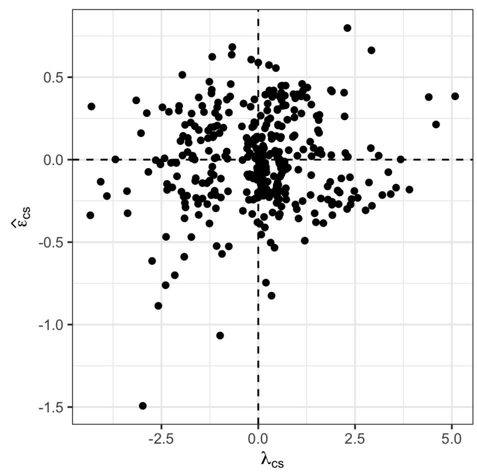
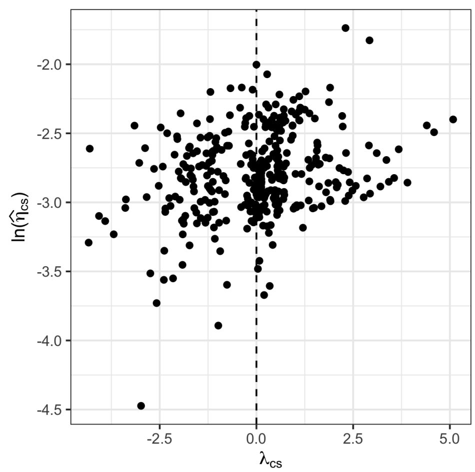


Figure 7: Here, we show the geo-spatial distribution of posterior means of log amenity preferences, $\widehat{\ln(\eta_{cs})}$, for all counties in our sample. Notice that all values of $\eta_{cs} \in (0, 1)$.



(a)



(b)

Figure 8: In panel (a) we present a scatterplot of λ_{cs} along with the posterior means of the idiosyncratic errors $\widehat{\epsilon}_{cs}$, and in panel (b) we present a scatterplot of λ_{cs} along with the posterior means of log amenity preferences, $\widehat{\ln(\eta_{cs})}$. The log efficiency index appears uncorrelated with both $\widehat{\epsilon}_{cs}$ and $\widehat{\ln(\eta_{cs})}$.

5.3.2 Checking for Regressor Endogeneity

A standard regression model contains a response (dependent) variable (often written on the left-hand side of an equation) and a vector of other, assumed independent variables that the response depends upon in some manner (model). Here, the variable $\frac{Rating_{cs}}{e_{cs}}$ depends non-linearly on relative prices, local solar insolation, local demographic factors, and a non-additive residual ϵ_{cs}, which in a standard regression would be assumed to be orthogonal to all but the dependent variable. We take no stance on what the model's dependent variable is, given general equilibrium linkages likely make both relative demand and relative prices endogenous. In this section we explore the covariance of the model variables with the error term to understand the degree to which endogeneity-bias may be present. If only one of the regression variables exhibits significant correlation with the error term, then we can conclude that no endogeneity bias is present. This is because, as in a canonical regression model, the covariance between the dependent variable and the error term is typically (but need not be) non-zero.

Table 4: p-Values From Boot-strapping Tests for Covariance Between $\widehat{\epsilon}_{cs}$ and Regression Variables

Panel (a): p-Values for Model with $\rho, \alpha \in (0,1)$ and Estimated Likelihood Variance $\widehat{\sigma}^2 = 0.104$

υ	$\ln\left(\frac{p^{solar}_{cs}}{p^{e}_{cs}}\right)$	$\frac{p^{h}_{cs}}{p^{e}_{cs}}$	$\ln\left(\frac{Rating_{cs}}{e_{cs}}\right)$	$\ln\overline{\phi}_{cs}$	$\ln pop\_density_{cs}$	$\ln GOP\_voting\_share_{cs}$	$\ln median\_income_{cs}$	$\ln heating\_days_{cs}$	$\ln cooling\_days_{cs}$	$\ln homeownership\_rate_{cs}$
0.05	0.687	0.009	0.100	0.992	0.271	0.661	0.914	0.305	0.585	0.989
0.1	0.959	0.018	0.197	1.000	0.512	0.946	1.000	0.573	0.895	1.000
0.25	1.000	0.047	0.473	1.000	0.921	1.000	1.000	0.953	1.000	1.000
0.5	1.000	0.090	0.798	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1	1.000	0.182	0.992	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Panel (b): p-Values for Model with $\rho \in (0,1)$, $\alpha < 0$ and Estimated Likelihood Variance $\widehat{\sigma}^2 = 0.110$

υ	$\ln\left(\frac{p^{solar}_{cs}}{p^{e}_{cs}}\right)$	$\frac{p^{h}_{cs}}{p^{e}_{cs}}$	$\ln\left(\frac{Rating_{cs}}{e_{cs}}\right)$	$\ln\overline{\phi}_{cs}$	$\ln pop\_density_{cs}$	$\ln GOP\_voting\_share_{cs}$	$\ln median\_income_{cs}$	$\ln heating\_days_{cs}$	$\ln cooling\_days_{cs}$	$\ln homeownership\_rate_{cs}$
0.05	0.697	0.010	0.103	0.993	0.274	0.673	0.915	0.314	0.586	0.992
0.1	0.962	0.021	0.203	1.000	0.519	0.950	0.999	0.585	0.900	1.000
0.25	1.000	0.047	0.484	1.000	0.931	1.000	1.000	0.957	1.000	1.000
0.5	1.000	0.093	0.800	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1	1.000	0.182	0.991	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Panel (c): p-Values for Model with $\rho < 0$, $\alpha \in (0,1)$ and Estimated Likelihood Variance $\widehat{\sigma}^2 = 5.872$

υ	$\ln\left(\frac{p^{solar}_{cs}}{p^{e}_{cs}}\right)$	$\frac{p^{h}_{cs}}{p^{e}_{cs}}$	$\ln\left(\frac{Rating_{cs}}{e_{cs}}\right)$	$\ln\overline{\phi}_{cs}$	$\ln pop\_density_{cs}$	$\ln GOP\_voting\_share_{cs}$	$\ln median\_income_{cs}$	$\ln heating\_days_{cs}$	$\ln cooling\_days_{cs}$	$\ln homeownership\_rate_{cs}$
0.05	1.000	0.055	0.586	1.000	0.976	1.000	1.000	0.990	1.000	1.000
0.1	1.000	0.110	0.899	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.25	1.000	0.272	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.5	1.000	0.522	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1	1.000	0.846	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Panel (d): p-Values for Model with $\rho, \alpha < 0$ and Estimated Likelihood Variance $\widehat{\sigma}^2 = 5.938$

υ	$\ln\left(\frac{p^{solar}_{cs}}{p^{e}_{cs}}\right)$	$\frac{p^{h}_{cs}}{p^{e}_{cs}}$	$\ln\left(\frac{Rating_{cs}}{e_{cs}}\right)$	$\ln\overline{\phi}_{cs}$	$\ln pop\_density_{cs}$	$\ln GOP\_voting\_share_{cs}$	$\ln median\_income_{cs}$	$\ln heating\_days_{cs}$	$\ln cooling\_days_{cs}$	$\ln homeownership\_rate_{cs}$
0.05	1.000	0.056	0.596	1.000	0.976	1.000	1.000	0.989	1.000	1.000
0.1	1.000	0.117	0.904	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.25	1.000	0.288	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.5	1.000	0.534	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1	1.000	0.854	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

To assess the degree to which our estimates may be biased by possible endogeneity of the regressors, we use a boot-strapping procedure to separately examine the covariance of $\widehat{\epsilon}_{cs}$ with each of the regression variables. Specifically, if X represents one of the regression variables, including one of the control variables, we construct $M = 10000$ random perturbations (re-orderings) of the vector X and compute the covariance of each random perturbation X_m with the actual vector of point estimates $\widehat{\epsilon}_{cs}$. This gives us a set of M test statistics for each regression variable X. In order to compute a p-value to test the null hypothesis that $\text{Cov}(X, \widehat{\epsilon}_{cs}) \neq 0$, we must specify a necessary threshold δ such that $|\text{Cov}(X, \widehat{\epsilon}_{cs})| < \delta$ implies that the covariance is effectively zero. This is because our test statistic is not necessarily distributed according to a standard Student's-t or Normal distribution. We thus choose several values for δ, each of which is some fraction υ of the estimated posterior mean likelihood variance $\widehat{\sigma}^2$. Specifically, we consider $\upsilon \in \{0.05, 0.1, 0.25, 0.5, 1\}$, so that δ is 5%, 10%, 25%, 50%, and 100% of the estimated likelihood variance. We can then use the δ thresholds to compute an associated p-value, which for variable X is just:

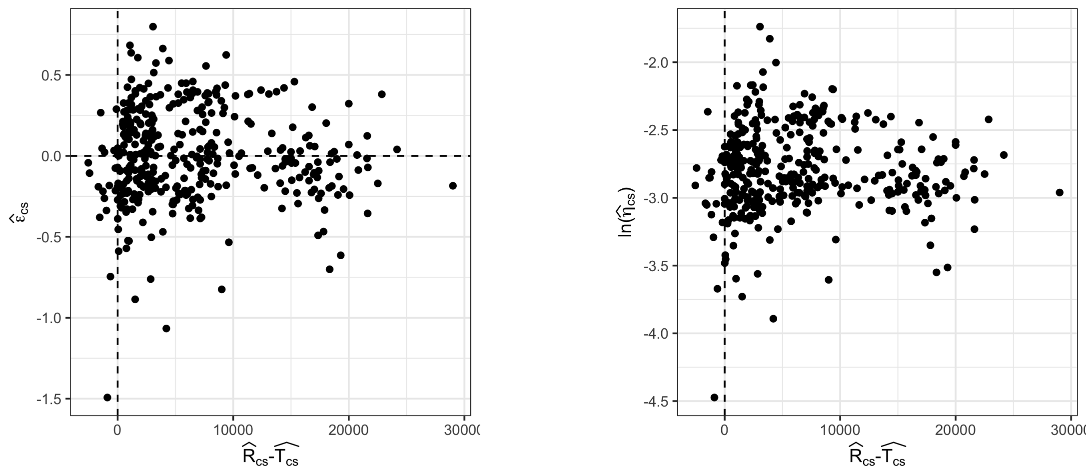
$$p_X = \frac{1}{M} \sum_m \mathbf{1}\Big\{ |\text{Cov}(X_m, \widehat{\epsilon}_{cs})| < \delta \Big\} \tag{10}$$

The null hypothesis that the covariances are non-zero is rejected if $1 - p_X$ is less than some pre-defined significance level.

The estimated p-values from the preferred models with $\rho \in (0, 1)$ are featured in panels (a) and (b) of Table 4. Note that the test we perform involves counting the amount of boot-strapped samples where $|\text{Cov}(X_m, \widehat{\epsilon}_{cs})| < \upsilon\, \widehat{\sigma}^2 \equiv \delta$. The results show that when we assume $\upsilon = 1$, thus assessing whether the covariances are greater than the estimated likelihood variance, we reject the null hypotheses that all but $\frac{p^h_{cs}}{p^e_{cs}}$ are endogenous at the 1% significance level (focussing specifically on Panels (a) and (b) of Table 4, which also feature relatively low likelihood variance $\widehat{\sigma}^2$ compared to the models where $\rho < 0$). Thus, covariances between regression variables and the likelihood error term are only a fraction of the marginal likelihood variance itself, suggesting endogeneity bias, if it exists, is not large in magnitude.

5.3.3 Correlation Between Errors, Amenity Preferences, and Local Subsidies

Figures 9a and 9b present scatterplots describing the relationship between net rebates, $\widehat{R}_{cs} - \widehat{T}_{cs}$, and both $\widehat{\epsilon}_{cs}$ and $\widehat{\ln \eta_{cs}}$, respectively. Not surprisingly, given the covariance tests undertaken in the previous section, we do not find significant correlations between rebate policies and either the idiosyncratic error or amenity preferences. Pearson's correlation coefficient describing the relationship between $\widehat{R}_{cs} - \widehat{T}_{cs}$ and $\widehat{\epsilon}_{cs}$ is -0.030 while the correlation coefficient describing the relationship between $\widehat{R}_{cs} - \widehat{T}_{cs}$ and $\widehat{\ln \eta_{cs}}$ is -0.001. These results suggest that there is no systematic bias being caused by rebate policies directly affecting solar price construction.



(a) (b)

Figure 9: Average county-level net rebates, $\widehat{R}_{cs} - \widehat{T}_{cs}$, also appear to be uncorrelated with idiosyncratic errors and log amenity preferences.

6 Counterfactual Thought Experiments

The main goal of this paper is to understand how the market share of electricity provided by residential solar is impacted by both local amenity preferences, including local environmental factors, and local subsidization policies. We thus engage in several counterfactual thought experiments, varying local characteristics, prices, and subsidies and then examining both how the geographic distribution of the efficiency index described in Section 3.5 changes and whether or not the share of electrical consumption that comes from residential solar increases.

We find that local climatic factors which cannot be altered by governmental policies have the biggest impact on residential solar uptake. Since many of the demographic control variables enter the log amenity weight negatively, varying them does not have a noticeable impact on solar panel uptake or the geo-spatial efficiency distribution. On the price side, targeted rebates are only effective at increasing the solar share of electrical consumption in high-insolation locales if housing and electricity are substitutes ($0 < \alpha < 1$). However, in this case the share also rises in low-insolation locales, biasing the geo-spatial efficiency distribution. When it is assumed that housing and electricity are complements ($\alpha < 0$), reducing the solar price through rebates actually has a *negative* impact on the solar share of electrical consumption in high-insolation regions. While solar installations increase as the price falls, electrical consumption increases faster. Thus, the rebate itself could have a *net-negative* impact in terms of environmental-policy goals to reduce carbon emissions if increases in e_{cs} are supplied by new construction of other, non-renewable energy sources. Thus, the effectiveness of net-rebates for solar installation costs in terms of actually increasing the renewable share of electrical consumption depends very importantly on consumer preferences. Our preferred model ($0 < \rho < 1$; $\alpha < 0$) thus predicts that one-size-fits-all rebate policies could actually backfire on policymakers whose aim is to increase the renewable share of electrical consumption.

6.1 Counterfactual Variation in Efficiency and the Solar Share Around Local Demographic Characteristics

How does variation in local demographics and climate impact solar demand? To answer this question, we simulate the solar share of electrical consumption $\frac{Rating_{cs}\overline{\phi}_{cs}}{e_{cs}}$, counterfactually holding the separate demographic and climatological factors that influence η_{cs} and $\overline{\phi}_{cs}$ fixed. We then compute posterior counterfactual mean log efficiency indices $\widetilde{\lambda}_{cs}$ and use these indices to understand how the geo-spatial distribution is affected by different demographic and climatological forces.[18]

Our simulation strategy is to compute summary statistics for the various factors that comprise $\ln \eta_{cs}$ and replace the local data observation for the variable of interest with those summary statistics. We perform similar analyses for $\ln \overline{\phi}_{cs}$ as well. The left-most columns in Tables 5 and 6 feature the specific counterfactuals we consider.[19]

Tables 5 and 6 show summary statistics for the simulated posterior distribution of $\widetilde{\lambda}_{cs}$ under the various counterfactual assumptions. Note that the means of these posteriors are not significantly different than zero, so that perturbing how demographics and climatological factors affect demand does not seem to introduce bias to the efficiency index. Focussing on the preferred models in Table 5 with emphasis on the bottom half of the table, notice that the standard deviation of the geo-spatial distribution of log efficiency, relative to data, is most reduced by fixing the log number of cooling days.[20] This is not

[18] Tildes are used to denote counterfactually-simulated variables.

[19] For example, in row 2 of Table 5, we show the summary statistics for the simulated posterior distribution of $\widetilde{\lambda}_{cs}$ under the assumption that every county is identical to the average county in terms of log GOP voting share. Thus, under this counterfactual, there is no nationwide heterogeneity in terms of log GOP voting share, yet all other variables are allowed to take on the county-specific values we observe in data.

[20] Compare the standard deviations (S.D.) in the last three rows of each sub-table to that in the first row.

Table 5: Counterfactual Distribution of Efficiency Index, $\widetilde{\lambda}_{cs}$, $\rho \in (0,1)$

$\alpha \in (0,1)$

Counterfactual	Mean	S.D.	2.5%	25%	50%	75%	97.5%
Data λ_{cs}	0.000	1.471	-2.954	-1.059	0.107	0.659	3.062
Mean Log GOP Voting Share	0.001	1.503	-2.850	-0.969	0.098	0.648	3.181
Mean Log $\overline{\phi}_{cs}$	0.001	1.492	-2.869	-0.970	0.069	0.657	3.262
10^{th} Percentile Log $\overline{\phi}_{cs}$	0.001	1.493	-2.869	-0.970	0.067	0.657	3.265
90^{th} Percentile Log $\overline{\phi}_{cs}$	0.001	1.492	-2.869	-0.970	0.071	0.657	3.252
Mean Log Homeownership	0.001	1.508	-3.205	-0.973	0.081	0.700	3.169
Mean Log Median Income	0.001	1.510	-3.015	-1.011	0.060	0.688	3.275
Mean Log Heating Days	0.001	1.601	-3.408	-1.026	0.050	0.780	3.363
Mean Log Cooling Days	0.001	1.050	-2.209	-0.354	0.007	0.338	2.628
10^{th} Percentile Log Cooling Days	0.001	1.073	-2.209	-0.415	-0.002	0.357	2.725
90^{th} Percentile Log Cooling Days	0.001	1.049	-2.209	-0.354	0.004	0.342	2.628

$\alpha < 0$

Counterfactual	Mean	S.D.	2.5%	25%	50%	75%	97.5%
Data λ_{cs}	0.000	1.471	-2.954	-1.059	0.107	0.659	3.062
Mean Log GOP Voting Share	0.001	1.523	-3.059	-0.851	-0.068	0.737	3.371
Mean Log $\overline{\phi}_{cs}$	0.001	1.508	-2.878	-0.783	-0.097	0.760	3.377
10^{th} Percentile Log $\overline{\phi}_{cs}$	0.001	1.508	-2.878	-0.783	-0.097	0.760	3.377
90^{th} Percentile Log $\overline{\phi}_{cs}$	0.001	1.508	-2.878	-0.783	-0.097	0.760	3.377
Mean Log Homeownership	0.001	1.522	-2.976	-0.836	-0.078	0.786	3.360
Mean Log Median Income	0.001	1.515	-2.866	-0.795	-0.110	0.769	3.381
Mean Log Heating Days	0.001	1.579	-3.104	-0.916	-0.033	0.859	3.387
Mean Log Cooling Days	0.001	1.294	-2.701	-0.552	-0.066	0.406	3.282
10^{th} Percentile Log Cooling Days	0.001	1.294	-2.701	-0.552	-0.066	0.406	3.282
90^{th} Percentile Log Cooling Days	0.001	1.294	-2.701	-0.552	-0.066	0.406	3.282

too surprising given that the log cooling days variable appears to have the strongest, most negative effect on preferences.

In Table 7 we compute the variance of the difference between the data log efficiency index and the counterfactual ones, $\text{Var}(\lambda_{cs} - \widetilde{\lambda}_{cs})$. The larger this statistic is, the greater the effect that the counterfactual perturbation has on altering the geo-spatial efficiency

Table 6: Counterfactual Distribution of Efficiency Index, $\widetilde{\lambda}_{cs}$, $\rho < 0$

$\alpha \in (0,1)$							
Counterfactual	Mean	S.D.	2.5%	25%	50%	75%	97.5%
Data λ_{cs}	0.000	1.471	-2.954	-1.059	0.107	0.659	3.062
Mean Log GOP Voting Share	0.001	1.391	-3.051	-0.730	0.159	0.749	2.651
Mean Log $\overline{\phi}_{cs}$	0.001	1.392	-3.181	-0.734	0.133	0.747	2.708
10^{th} Percentile Log $\overline{\phi}_{cs}$	0.001	1.389	-3.181	-0.725	0.134	0.753	2.712
90^{th} Percentile Log $\overline{\phi}_{cs}$	0.001	1.397	-3.181	-0.733	0.136	0.747	2.695
Mean Log Homeownership	0.001	1.397	-3.072	-0.677	0.128	0.761	2.691
Mean Log Median Income	0.001	1.408	-3.022	-0.790	0.149	0.761	2.718
Mean Log Heating Days	0.001	1.441	-3.125	-0.885	0.207	0.780	2.727
Mean Log Cooling Days	0.001	1.224	-2.909	-0.502	-0.005	0.661	2.263
10^{th} Percentile Log Cooling Days	0.001	1.193	-2.909	-0.461	0.0001	0.666	2.142
90^{th} Percentile Log Cooling Days	0.001	1.249	-2.909	-0.518	-0.004	0.666	2.305

$\alpha < 0$							
Counterfactual	Mean	S.D.	2.5%	25%	50%	75%	97.5%
Data λ_{cs}	0.000	1.471	-2.954	-1.059	0.107	0.659	3.062
Mean Log GOP Voting Share	0.001	1.361	-3.121	-0.597	0.138	0.748	2.636
Mean Log $\overline{\phi}_{cs}$	0.001	1.363	-3.250	-0.645	0.135	0.737	2.658
10^{th} Percentile Log $\overline{\phi}_{cs}$	0.001	1.358	-3.250	-0.636	0.142	0.736	2.648
90^{th} Percentile Log $\overline{\phi}_{cs}$	0.001	1.368	-3.250	-0.657	0.134	0.748	2.646
Mean Log Homeownership	0.001	1.369	-3.227	-0.587	0.126	0.748	2.662
Mean Log Median Income	0.001	1.380	-3.272	-0.673	0.148	0.739	2.690
Mean Log Heating Days	0.001	1.406	-3.182	-0.732	0.195	0.761	2.707
Mean Log Cooling Days	0.001	1.206	-3.087	-0.500	0.012	0.646	2.268
10^{th} Percentile Log Cooling Days	0.001	1.181	-2.994	-0.477	0.035	0.629	2.183
90^{th} Percentile Log Cooling Days	0.001	1.232	-3.063	-0.524	0.008	0.635	2.284

distribution. Again, changing log cooling days appears to most change the geo-spatial efficiency distribution, relative to what we estimate from data. The largest effect comes from fixing the log cooling days at the 90^{th} percentile, which is approximately equivalent to making every county have climatological conditions, and thus the need for air conditioning, like that of Yuma, Arizona. Further, we find that if the entire country were both

Table 7: Variance of Re-distribution of Efficiency Index, $\text{Var}(\lambda_{cs} - \widetilde{\lambda}_{cs})$

	Model			
$\widetilde{\lambda}_{cs}$	$\rho, \alpha \in (0,1)$	$\rho \in (0,1), \alpha < 0$	$\rho < 0, \alpha \in (0,1)$	$\rho, \alpha < 0$
Mean Log GOP Voting Share	0.194	0.688	0.305	0.467
Mean Log $\overline{\phi}_{cs}$	0.049	0.761	0.343	0.521
10^{th} Percentile Log $\overline{\phi}_{cs}$	0.050	0.761	0.351	0.536
90^{th} Percentile Log $\overline{\phi}_{cs}$	0.049	0.761	0.330	0.506
Mean Log Homeownership	0.112	0.732	0.319	0.483
Mean Log Median Income	0.047	0.797	0.321	0.502
Mean Log Heating Days	0.475	0.759	0.312	0.478
Mean Log Cooling Days	1.268	1.247	0.582	0.725
10^{th} Percentile Log Cooling Days	1.222	1.247	0.661	0.824
90^{th} Percentile Log Cooling Days	1.274	1.247	0.524	0.632

as sunny and as hot as Yuma, the solar share would actually fall across the board.

Qualitatively, the effect of demand for air conditioning (for which cooling days is a proxy) is to enhance the inefficiency of the geo-spatial distribution of panels by essentially applying downward pressure on relative demand in sunny but hot places. The results here thus suggest that while political preferences and other local demographic factors are mildly important in affecting demand for residential solar, basic climatological factors seem to play the most significant role.

6.2 How Does Solar Price Variation Affect Adoption and Efficiency?

In this section we examine how the share of electrical consumption devoted to residential solar changes as solar prices vary and how such changes impact the geo-spatial efficiency distribution. Varying p_{cs}^{solar} acts as a proxy for varying underlying rebates and taxes. If prices fall, we can think of this as the local government offering more rebates or lower sales taxes, while if prices rise, rebates fall and taxes rise. Our findings suggest that whether rebates will increase the solar share of electrical consumption in high-insolation places depends on if housing and electricity are substitutes ($0 < \alpha < 1$) or complements ($\alpha < 0$). Reducing prices substantially across the country (either through federal rebates or other subsidies) will only have a positive impact on the solar share of electrical consumption if housing and electricity are substitutes. If they are complements (our theoretically preferred model), solar installations rise, since solar demand is a normal good, but electrical consumption rises faster in response to a solar price decline, so that in almost all localities across the country, the residential solar share of electrical consumption falls. In terms of geo-spatial efficiency, if $\alpha \in (0,1)$, decreasing solar prices leads to the distribution of $\widetilde{\lambda}_{cs}$ becoming biased towards counties with relative excess supply. If $\alpha < 0$, decreasing solar prices may encourage more efficient solar adoption, despite the solar share of electrical consumption falling. The effect, in the latter case however, is very slight.

To illustrate how the residential solar share of electrical consumption responds to price variation, we simulate demand assuming that all counties face the minimum observed price, p_{cs}^{solar}, in our data.[21] In Figures 10 and 11 we plot the posterior median change in solar share by county, relative to that observed in the data, under the counterfactual pricing assumption. For these plots, we present the posterior median change due to significant left-skewness in the counterfactual posterior distributions of $\widetilde{\frac{solar_{cs}}{e_{cs}}}$. The value shown in the heat map is thus the posterior median of the difference $\widetilde{\frac{solar_{cs}}{e_{cs}}} - \frac{solar_{cs}}{e_{cs}}$.

When the price of solar in all counties is equivalent to the minimum price observed in our data, over 99% of counties see a rise in solar share if $\alpha \in (0,1)$, but less than 1% see a rise if $\alpha < 0$. Recall, though, since solar is a normal good, solar installations do increase, even if the share does not. This then implies that overall electrical consumption must be increasing faster. Focussing on Figure 11, notice that when $\alpha < 0$ the smallest declines in solar share are geographically concentrated in cooler environs with lower average daily insolation, such as the Pacific Northwest and Northeast. Notice that solar shares rise across the board when $\alpha \in (0,1)$ as shown in Figure 10. Thus, depending on whether we believe housing and electrical consumption are substitutes or complements, rebates that reduce the price of solar may be wildly effective at increasing the solar share across the board (substitutes) or very ineffective except in cooler environments (complements).

Since prices are decreasing in rebates, this analysis provides a proxy for how increasing rebates in various counties would affect residential solar uptake. Under our preferred model ($\alpha < 0$), little is changed by rebates in regions that are both sunnier with higher average insolation, and have comparatively high shares of households with installed capacity (Figure 2). In such regions, which are also associated with high cooling degree days weighing negatively on solar amenity preferences, solar share price responsiveness appears to be very inelastic, while it is slightly less so in regions with lower cooling degree days. This counterfactual thus helps underscore our conclusion that cooling degree days

[21]The minimum price is that for Jefferson County, Arkansas.

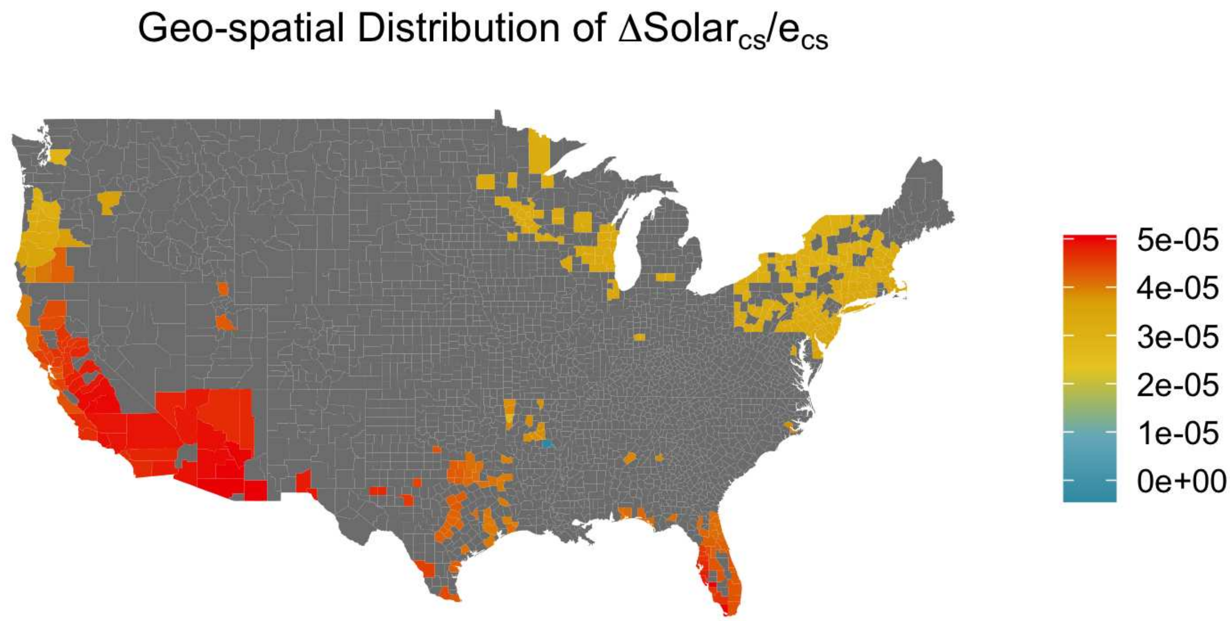


Figure 10: $\rho, \alpha \in (0, 1)$; This is the median change in the residential solar share of electricity by county if all counties faced the minimum observed price. Note that the solar share increases in most counties when housing and electricity are substitutes.

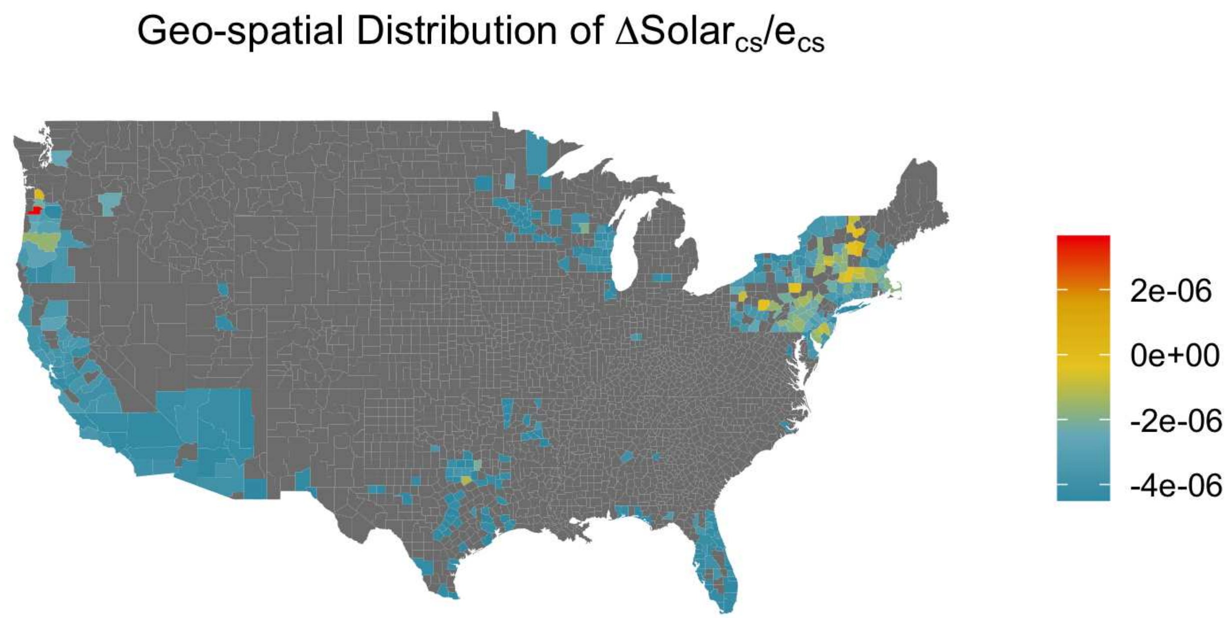


Figure 11: $\rho \in (0, 1), \alpha < 0$; Here, again, we present the median change in residential solar share. However, when housing and electricity are complements, as they are here, the solar share falls in most counties except a handful of counties in cloudier, more Northern regions.

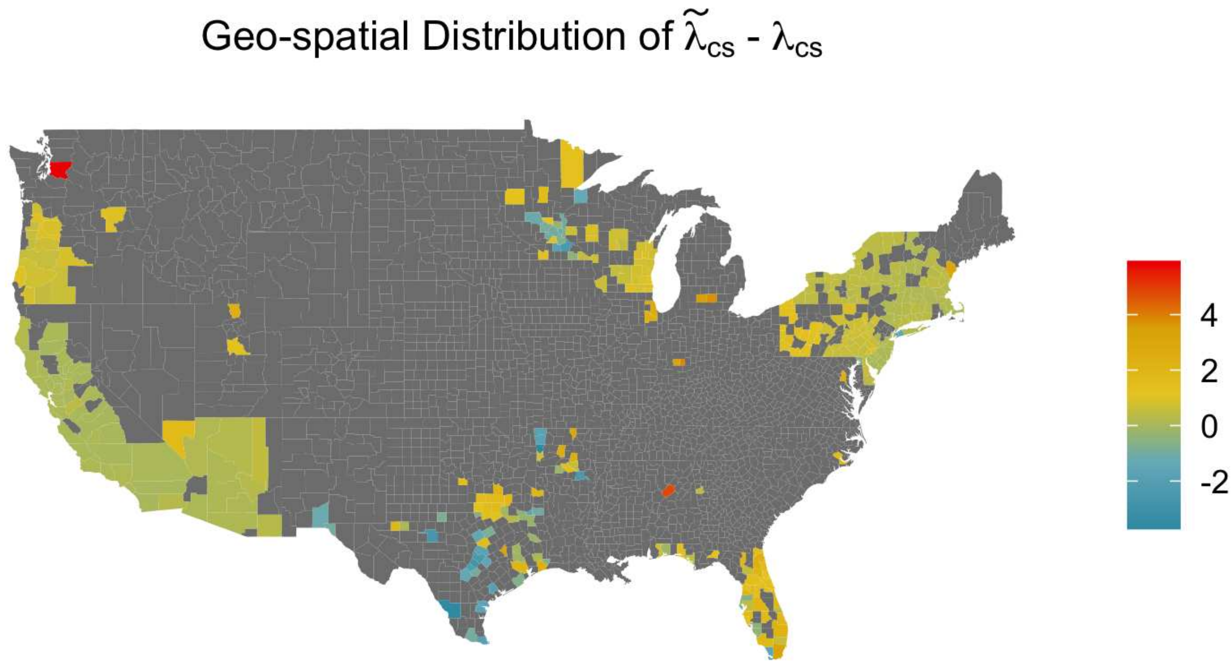


Figure 12: $\rho, \alpha \in (0, 1)$; Here, we show the median change in relative efficiency under the minimum-pricing counterfactual when housing and electricity are substitutes.

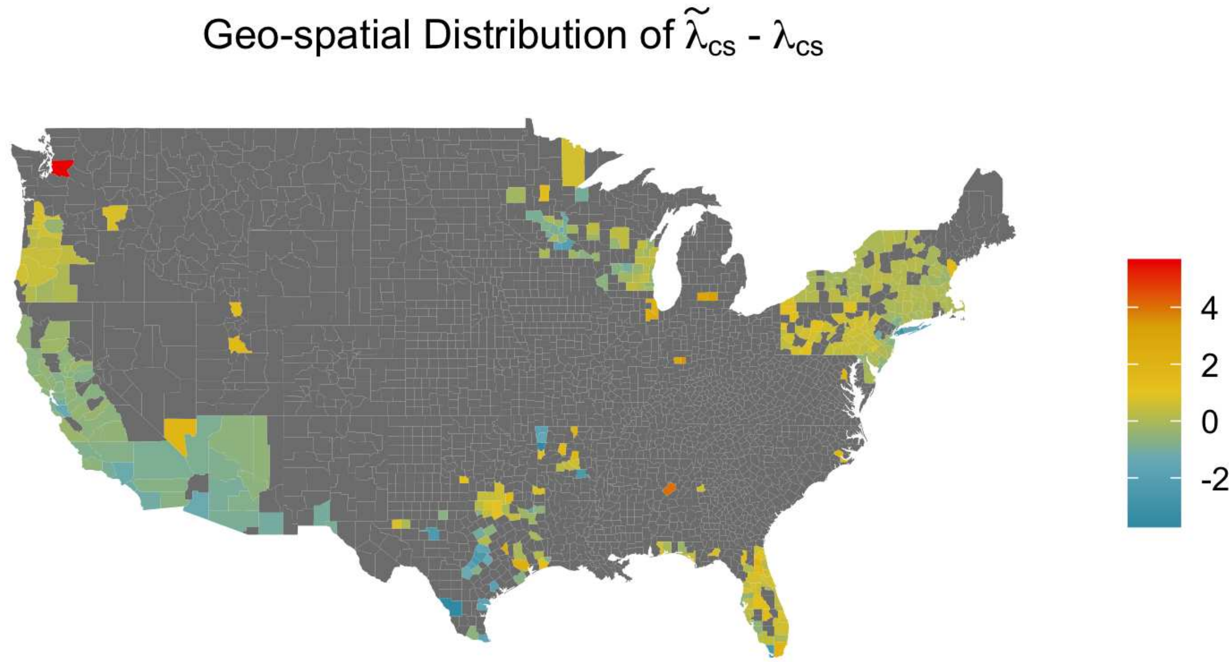


Figure 13: $\rho \in (0, 1), \alpha < 0$; This is the same plot as above, except we now show the median change in relative efficiency when housing and electricity are assumed to be complements.

that proxy as demand for air conditioning have a strong impact on the rate of local solar adoption. Thus, panel installation rebates appear to be relatively ineffective at increasing the solar share in environments featuring high-insolation (but also high cooling-degree days).

Figures 12 and 13 demonstrate how the efficiency distributions change in response to price reductions under our different parameterizations of α. If housing and electricity are substitutes, then despite the solar share rising across the board it appears to rise relatively faster in low-insolation environments, causing the efficiency distribution to be both biased —$\mathbb{E}(\widetilde{\lambda}_{cs}) = 0.462$ — and less efficient — $\mathrm{Var}(\widetilde{\lambda}_{cs}) = 2.713$ versus $\mathrm{Var}(\lambda_{cs}) = 2.164$ in the data. In this case, rebates may be effective at increasing share and reducing overall reliance on non-solar electrical consumption, but hotter places will still have relative under-capacity compared to cooler places which already start with relative excess capacity. When housing and electricity are complements, as is theoretically preferred, $\mathbb{E}(\widetilde{\lambda}_{cs}) = 0.001$ and $\mathrm{Var}(\widetilde{\lambda}_{cs}) = 2.647$, so the distribution of panels still becomes less geo-spatially efficient but also remains unbiased. Despite the distribution being slightly unchanged, notice that some locales in Texas and Florida with relative under-supply do receive a slight boost to their efficiency ranking under the minimum-pricing counterfactual. The areas that appear to gain relative supply (brighter yellow, orange, and red) already have excess supply (Pacific Northwest and Northeast), though the Dallas Fort-Worth Metroplex in Texas and most of Florida are notable exceptions. Many sunnier locales in California and Arizona, meanwhile, fall on the efficiency index so that there is increasing relative under-utilization of solar potential in the Western U.S., even as solar installations rise. This is a direct consequence of wealth effects, governed by $\alpha < 0$, driving up electrical consumption faster than solar adoption, thus pushing down solar's share. Depending on assumptions regarding α, using rebates to encourage adoption may have mixed results and actually exacerbate the geo-spatial efficiency of the aggregate panel distribution.

7 Conclusion

We have shown, using model-free empirical evidence, that the aggregate geo-spatial distribution of residential solar panels in the U.S. is inefficient, absent considering transmission constraints between different regions of the country. Through the lens of a structural model, we demonstrate that if housing and electricity are indeed complements, local rebates encouraging households to install residential solar panels may backfire and have a net-negative impact on the residential solar share of electrical consumption. This is because we estimate that the local demand for air conditioning, as proxied by cooling degree days, weighs negatively on the solar share. Given these considerations, if policymakers seek to enact subsides and rebates that increase the residential solar share of electrical consumption, policies which help consumers pay for still nascent but expensive battery-storage technologies may be more effective at increasing the residential solar share in the future.

A Posterior Estimates Assuming Maximum ITC Uptake

In this appendix we assume that all consumers take advantage of the federal government's 30% tax credit against the cost of panel purchase and installation. Thus, the price of solar panels p_{cs}^{solar} is constructed to account for this tax credit. Broadly speaking, changing our pricing assumptions yields no significant changes to posterior distribution estimation results.

Table A.1: Model Selection Assessments, 30% ITC Uptake

Model	$elppd$ Difference[a]	S.E. Difference[b]
$\rho, \alpha \in (0,1)$	—	—
$\rho \in (0,1), \alpha < 0$	-15.7	3.8
$\rho < 0, \alpha \in (0,1)$	-815.8	19.7
$\rho, \alpha < 0$	-817.4	19.6

[a] This is the difference in the expected log-posterior predictive density relative to the model with the highest $elppd$ value — $\rho, \alpha \in (0,1)$. As long as the absolute value of the $elppd$ difference is several times the standard error of the differences, we can be confident that the baseline model provides a better fit than its alternatives (Vehtari, Gelman, and Gabry 2017).

[b] This is the standard error of the posterior differences across HMC sample draws.

In Table A.1 we see that model selection, under this alternative pricing assumption, proceeds in the same manner as in the main text. Models where $\rho \in (0,1)$ are strongly preferred to those where $\rho < 0$ based on the $elppd$ criterion of Vehtari, Gelman, and Gabry (2017). Further, the magnitudes of differences in $elppd$ relative to the highest ranking model are very similar to those featured in Table 1 in the main text. As in the main text, under this alternative pricing assumption we still prefer the model where $\rho \in (0,1)$ and $\alpha < 0$ for theoretical reasons despite it ranking second in terms of $elppd$ differentials. As can be seen in the top half of Table A.2, values for γ are very close to 1, just as in the main text, and α skews toward 1, so that housing and electrical consumption are strong substitutes and consumers do not put much weight on their electricity choices. This seems counterintuitive for the same reasons we previously argued.

To see that posterior estimates are similar to those in the main text, compare the means of the structural parameters in both Tables A.2 and A.3 to those in Tables 2 and 3. ρ is little changed compared to estimates featured in the main text, and in fact has an identical mean under both estimation procedures when $\rho < 0$. Variations in α, γ, and σ^2 between the different specifications are also similar to those in the main text, as are the signs of the components of $\boldsymbol{\beta}$. We conclude that accounting for the possibility of consumer uptake of federal tax credits has no significant impact on model inference and estimation outcomes.

Table A.2: Posterior Distribution Estimates, $\rho \in (0,1)$, 30% ITC Uptake

$\alpha \in (0,1)$

Parameter	$\widehat{R}$	Mean	S.D.	2.5%	25%	50%	75%	97.5%
ρ	1.000	0.969	0.007	0.956	0.964	0.969	0.973	0.982
α	1.000	0.809	0.081	0.607	0.770	0.823	0.867	0.921
γ	1.001	0.973	0.015	0.937	0.970	0.977	0.981	0.987
σ^2	1.000	0.098	0.007	0.084	0.093	0.097	0.102	0.113
β_1	1.000	0.020	0.023	-0.026	0.005	0.020	0.036	0.065
β_2	1.000	0.078	0.063	-0.046	0.036	0.078	0.121	0.202
β_3	1.000	-0.098	0.070	-0.235	-0.145	-0.097	-0.050	0.041
β_4	1.000	-0.055	0.037	-0.127	-0.080	-0.056	-0.030	0.016
β_5	1.000	-0.437	0.072	-0.578	-0.485	-0.437	-0.387	-0.295
β_6	1.000	-0.269	0.144	-0.553	-0.365	-0.267	-0.173	0.013
$V(\mathcal{P})$	1.000	245.927	2.350	240.439	244.573	246.264	247.661	249.476

$\alpha < 0$

Parameter	$\widehat{R}$	Mean	S.D.	2.5%	25%	50%	75%	97.5%
ρ	1.000	0.963	0.007	0.950	0.958	0.963	0.967	0.976
α	1.001	-0.147	0.098	-0.397	-0.189	-0.125	-0.080	-0.031
γ	1.001	0.324	0.110	0.104	0.249	0.327	0.400	0.532
σ^2	1.000	0.104	0.008	0.090	0.099	0.104	0.109	0.120
β_1	1.000	0.009	0.024	-0.039	-0.007	0.009	0.025	0.056
β_2	1.000	0.090	0.065	-0.037	0.047	0.091	0.134	0.218
β_3	1.001	0.011	0.082	-0.149	-0.045	0.011	0.066	0.175
β_4	1.001	-0.037	0.037	-0.111	-0.063	-0.037	-0.012	0.035
β_5	1.001	-0.369	0.073	-0.513	-0.419	-0.369	-0.320	-0.227
β_6	1.000	-0.229	0.146	-0.515	-0.327	-0.229	-0.131	0.056
$V(\mathcal{P})$	1.000	233.725	2.289	228.423	232.438	234.048	235.381	237.192

Table A.3: Posterior Distribution Estimates, $\rho < 0$, 30% ITC Uptake

$\alpha \in (0,1)$								
Parameter	$\widehat{R}$	Mean	S.D.	2.5%	25%	50%	75%	97.5%
ρ	1.000	-0.011	0.005	-0.023	-0.014	-0.010	-0.008	-0.004
α	1.000	0.568	0.266	0.047	0.357	0.614	0.800	0.957
γ	1.001	0.790	0.253	0.104	0.697	0.919	0.964	0.995
σ^2	1.000	6.064	0.452	5.245	5.745	6.040	6.356	7.019
$V(\mathcal{P})$	1.001	-543.339	2.660	-549.175	-544.971	-543.137	-541.483	-538.684
β_1	1.000	-0.203	0.180	-0.559	-0.324	-0.204	-0.081	0.152
β_2	1.000	-1.591	0.481	-2.530	-1.919	-1.591	-1.263	-0.655
β_3	1.000	0.492	0.541	-0.593	0.131	0.493	0.860	1.547
β_4	1.000	-0.194	0.290	-0.776	-0.386	-0.192	0.002	0.367
β_5	1.000	-2.055	0.554	-3.154	-2.421	-2.058	-1.690	-0.957
β_6	1.000	3.213	1.123	1.014	2.460	3.219	3.963	5.407

$\alpha < 0$								
Parameter	$\widehat{R}$	Mean	S.D.	2.5%	25%	50%	75%	97.5%
ρ	1.000	-0.011	0.005	-0.022	-0.014	-0.010	-0.007	-0.004
α	1.001	-0.498	0.562	-1.897	-0.597	-0.341	-0.190	-0.062
γ	1.000	0.660	0.314	0.034	0.405	0.752	0.951	1.000
σ^2	1.000	6.125	0.459	5.282	5.803	6.102	6.420	7.102
$V(\mathcal{P})$	1.001	-544.424	2.376	-550.101	-545.740	-544.075	-542.689	-540.831
β_1	1.000	-0.249	0.179	-0.597	-0.370	-0.249	-0.127	0.097
β_2	1.000	-1.553	0.484	-2.503	-1.880	-1.550	-1.230	-0.609
β_3	1.000	0.231	0.545	-0.850	-0.128	0.230	0.589	1.305
β_4	1.000	-0.071	0.272	-0.616	-0.252	-0.071	0.109	0.461
β_5	1.000	-1.952	0.546	-3.037	-2.312	-1.955	-1.588	-0.870
β_6	1.000	3.101	1.109	0.935	2.349	3.095	3.853	5.281

References

Bakkensen, Laura, and Paul Schuler. 2020. "A preference for power: Willingness to pay for energy reliability versus fuel type in Vietnam". *Energy Policy* 144:111696. (Cit. on p. 3).

Barbose, Galen, et al. 2019. "Tracking the Sun: Pricing and Design Trends for Distributed Photovoltaic Systems in the United States-2019 Edition". (Cit. on p. 1).

Bollinger, Bryan, and Kenneth Gillingham. 2012. "Peer effects in the diffusion of solar photovoltaic panels". *Marketing Science* 31 (6): 900–912. (Cit. on p. 3).

Borenstein, Severin. 2017. "Private net benefits of residential solar PV: The role of electricity tariffs, tax incentives, and rebates". *Journal of the Association of Environmental and Resource Economists* 4 (S1): S85–S122. (Cit. on p. 3).

Borenstein, Severin, and Lucas W Davis. 2016. "The distributional effects of US clean energy tax credits". *Tax Policy and the Economy* 30 (1): 191–234. (Cit. on p. 3).

Callaway, Duncan S, Meredith Fowlie, and Gavin McCormick. 2018. "Location, location, location: The variable value of renewable energy and demand-side efficiency resources". *Journal of the Association of Environmental and Resource Economists* 5 (1): 39–75. (Cit. on p. 3).

EIA. 2020. "Monthly energy review". *DOE/EIA-0035 (2020/9). Office of Energy Statistics, US Department of Energy Washington*. (Cit. on p. 1).

Flowers, Mallory E, et al. 2016. "Climate impacts on the cost of solar energy". *Energy Policy* 94:264–273. (Cit. on p. 10).

Freeman, Janine M, et al. 2018. *System Advisor Model (SAM) General Description (Version 2017.9. 5)*. Tech. rep. National Renewable Energy Lab.(NREL), Golden, CO (United States). (Cit. on p. 2).

Gillingham, Kenneth, and James H Stock. 2018. "The cost of reducing greenhouse gas emissions". *Journal of Economic Perspectives* 32 (4): 53–72. (Cit. on p. 3).

Gorman, William M. 1959. "Separable utility and aggregation". *Econometrica: Journal of the Econometric Society*: 469–481. (Cit. on p. 17).

Heal, Geoffrey M. 2020. "Economic Aspects of the Energy Transition". *NBER working paper*, no. w27766. (Cit. on p. 3).

Heng, Yan, et al. 2020. "The heterogeneous preferences for solar energy policies among US households". *Energy Policy* 137:111187. (Cit. on p. 3).

Hughes, Jonathan E, and Molly Podolefsky. 2015. "Getting green with solar subsidies: evidence from the California solar initiative". *Journal of the Association of Environmental and Resource Economists* 2 (2): 235–275. (Cit. on p. 3).

Imelda, Imelda, Matthias Fripp, and Michael Roberts. 2018. *Variable pricing and the cost of renewable energy*. Tech. rep. University of Hawaii Economic Research Organization, University of Hawaii at … (Cit. on p. 3).

Jordan, Dirk C, and Sarah R Kurtz. 2013. "Photovoltaic degradation rates—an analytical review". *Progress in photovoltaics: Research and Applications* 21 (1): 12–29. (Cit. on p. 10).

Neal, Radford. 2011. "MCMC using Hamiltonian dynamics". Chap. 5 in *Handbook of Markov Chain Monte Carlo*, ed. by Steve Brooks et al. Chapman & Hall. (Cit. on p. 18).

Nomura, Noboru, and Makoto Akai. 2004. "Willingness to pay for green electricity in Japan as estimated through contingent valuation method". *Applied Energy* 78 (4): 453–463. (Cit. on p. 3).

Sengupta, Manajit, et al. 2018. "The national solar radiation data base (NSRDB)". *Renewable and Sustainable Energy Reviews* 89:51–60. (Cit. on p. 2).

Sexton, Steven E, et al. 2018. *Heterogeneous environmental and grid benefits from rooftop solar and the costs of inefficient siting decisions*. Tech. rep. National Bureau of Economic Research. (Cit. on p. 3).

Vehtari, Aki, Andrew Gelman, and Jonah Gabry. 2017. "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC". *Statistics and Computing* 27:1413–1432. (Cit. on pp. 18, 21, 22, 41).

Vehtari, Aki, et al. 2019. "Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC". *arXiv:1903.08008*. (Cit. on p. 22).

Wiser, Ryan, and Dev Millstein. 2020. "Evaluating the economic return to public wind energy research and development in the United States". *Applied Energy* 261:114449. (Cit. on p. 2).

Wolak, Frank A. 2018. *The Evidence from California on the Economic Impact of Inefficient Distribution Network Pricing*. Tech. rep. National Bureau of Economic Research. (Cit. on p. 3).

Yoo, Seung-Hoon, and So-Yoon Kwak. 2009. "Willingness to pay for green electricity in Korea: A contingent valuation study". *Energy policy* 37 (12): 5408–5416. (Cit. on p. 3).

Yu, Jiafan, et al. 2018. "DeepSolar: A machine learning framework to efficiently construct a solar deployment database in the United States". *Joule* 2 (12): 2605–2617. (Cit. on p. 4).