Optimal Solar Subsidies*

Mark Colas
University of Oregon

Emmett Saulnier
University of Oregon

April 26, 2023

Abstract

We study the spatial misallocation resulting from subsidies for residential solar panels in the US and quantify the associated environmental costs. We build a structural model of solar panel demand and electricity production across the country and estimate the model by combining 1) remotely sensed data on residential solar panels, 2) power-plant-level data on hourly production and emissions, and 3) a state-of-the-art air pollution model. The current subsidies lead to severe spatial misallocation. The optimal cost-neutral reform generates environmental benefits equal to those of a 6-11% increase in the productivity of residential solar panels nationally.

JEL Classification: H21, H23, Q42, Q48

Keywords: rooftop solar, optimal taxation, renewable energy

*mcolas@uoregon.edu and emmetts@uoregon.edu. We would like to thank Severin Borenstein, Jon Davis, David Evans, Stephie Fried, Simon Fuchs, Joshua Graff Zivin, Ashley Langer, Keaton Miller, John M. Morehouse, Nick Muller, Ed Rubin, Chris Timmins, Woan Foong Wong, and Eric Zou. This work benefited from access to the University of Oregon high performance computer, Talapas.

1 Introduction

There are considerable differences across the United States in the extent to which states subsidize solar panels and in solar panel installation rates. The environmental benefits of solar panel installation vary geographically as well. All else equal, the environmental benefits of installing solar panels are likely to be the largest in sunny areas and areas where high-polluting power plants would otherwise produce electricity. Economists have argued that the current system of subsidies does not properly reflect this spatial heterogeneity in the environmental benefits (Sexton et al., 2021).[1]

To what extent does this system of subsidies lead to a misallocation of residential solar panels across space? What would be the environmental benefits of switching to the optimal system of subsidies? This paper quantifies the optimal solar subsidies through the lens of a structural model of solar panel installation and electricity production. Heterogeneous households across the country choose the number of solar panels to install, accounting for the installation cost, the lifetime value of the electricity produced and subsidies received, and the nonpecuniary costs and benefits of installation. Households can also purchase electricity produced by a system of power plants. Individual power plants vary in the extent to which their production leads to environmental damages, their production capacity, and their location, which dictates how the grid transmits the plant's electricity across geographic regions.

Residential solar installations reduce environmental damage by decreasing fossil-fuel power plants' electricity production. Therefore, panels installed in areas with more sunlight have greater environmental benefits because they lead to larger decreases in electricity produced by these plants. The environmental benefits of solar panel installations also vary geographically because of differences in the distribution of technology employed by power plants across space. Panels installed in areas where environmentally unfriendly plants would otherwise produce electricity will be more beneficial than panels installed in areas with cleaner plants.

These environmental benefits are not internalized by the household, thus suggesting a role for government intervention.[2] The primary tool currently employed by US policymakers to deal with this externality is a system of federal and state subsidies for solar panels. We use the model to solve for the optimal subsidies subject to an exogenous government budget constraint and quantify the benefits of switching from the current system of subsidies to the optimal subsidies.[3] Doing so requires understanding how solar panel installation rates and

[1]See Holland et al. (2016), for example, for a similar argument for the benefits of geographically targeted subsidies for electric vehicles.

[2]Environmental externalities are the only source of inefficiency in our model. Thus, we abstract away from inefficiencies arising from market power, information frictions, and borrowing constraints.

[3]We focus on the optimal choice of solar subsidies and do not allow for other types of government intervention, such as pricing the externality via a carbon tax.

the damages associated with electricity production would change in response to alternative subsidy schemes. Therefore, our approach is to estimate a quantitative version of our model to calculate the optimal policies and the associated benefits.

Our primary data source is the DeepSolar Project (Yu et al., 2018), a dataset of the universe of residential solar panel installations in the contiguous US. Deepsolar uses a machine-learning framework to identify solar panel installations from satellite imagery. We supplement these data with data from Google Project Sunroof, another satellite-imagery-based dataset that provides information on solar irradiance across the US and on the number and size of rooftops suitable for solar panel installations. Combined, these two datasets provide the distribution and size of solar panel installations as well as solar irradiance and space suitable for solar panels across the US. We utilize these data to estimate the household component of the model via generalized method of moments, where we target the distribution and size of solar panel installations across census tracts in the US. Though sparsely parameterized, our household installation model matches the spatial distribution of installations well. We also show that our estimates are consistent with quasi-experimental evidence on the responsiveness of installations with respect to solar rebates (Hughes and Podolefsky, 2015; Crago and Chernyakhovskiy, 2017).

For data on power plants, we use Open Grid Emissions (OGE) data, which provides hourly production and emissions data covering nearly every power plant in the United States. We use these data to estimate plant-specific policy functions which map electricity demand and renewable production across the country to plant-level electricity production and emissions. We show that this power plant model matches the data's temporal and spatial distribution of electricity generation. Finally, we translate these emissions into environmental damages using AP3, a state-of-the-art integrated air pollution model.

We then use the estimated structural model to solve for the optimal cost-neutral subsidy reforms and quantify the misallocation caused by the current subsidy system. Our main result is that the current subsidy system leads to a severe misallocation of solar panels across space. Consider Washington, for example, a state where current subsidies are high even though sunlight is low and households receive marginal energy from relatively environmentally friendly power plants. We find that solar panels in Washington are over-subsidized by 80% relative to the optimal subsidy system, leading to 160% greater installations than optimal.[4] Decreasing subsidies in Washington would lead to large decreases in fiscal costs with small decreases in environmental benefits. On the other hand, in West Virginia, where current subsidies are low and the environmental benefits of solar installations are high, we find that

[4]That is, the total value of subsidies an average household would receive is 80% higher under the current subsidies than the optimal subsidies. Total installations in Washington are 2.6 times higher under the current subsidies than installations under the optimal subsidies.

panels are under-subsidized by nearly 70%, leading to installations that are 80% lower than optimal. More generally, panels are under-allocated by roughly 40% in the Midwest and South and over-allocated by 50% in the Northeast.[5]

We find that the misallocation caused by the current system of subsidies leads to substantial environmental costs. Switching from the current subsidies to the welfare-maximizing subsidies leads to a 6% increase in aggregate environmental benefits—environmental damages decrease by approximately the same amount as a 6% increase in the productivity of every rooftop solar panel in the US. Switching to subsidies set by a planner aiming to minimize environmental damages rather than maximize welfare would lead to an 11% increase in aggregate environmental benefits.

Finally, we compare the effects of marginal subsidy changes around the current system of subsidies. We find large differences in the cost-effectiveness of subsidy increases across states. For example, the environmental damages offset per dollar of government funds associated with subsidy increases in West Virginia are six times greater than the damages offset per dollar of subsidy increases in Washington. These results highlight that cost-neutral shifts around the current system of subsidies could lead to substantial decreases in environmental damages.

The remainder of the paper consists of various extensions and robustness checks. We analyze the sensitivity of our results to 1) alternative specifications of household preferences, 2) accounting for line losses in transmitting electricity from plants to homes, 3) the introduction of improved electricity storage technology, and 4) changes in utility-scale renewable electricity production. We find that the optimal system of subsidies remains qualitatively the same across these specifications.

To the best of our knowledge, ours is the first paper to quantify the environmental consequences of residential solar subsidies. In this respect, our paper is related to recent papers using quantitative approaches to measure the environmental consequences of subsidizing renewable energy or environmentally friendly goods (e.g., Liski and Vehviläinen (2020), Shapiro (2021), Holland, Mansur, and Yates (2021), Jacobsen et al. (2022), Holland, Mansur, and Yates (2022), Arkolakis and Walsh (2022)).[6]

This paper is also related to a literature estimating the extent to which the marginal benefits of renewable energy investments vary geographically (e.g., Holland and Mansur (2008), Graff Zivin, Kotchen, and Mansur (2014), Holland et al. (2016), Callaway, Fowlie, and Mc-

[5] In Section 6.3, we calculate the optimal system of subsidies when subsidies are allowed to vary nonparametrically by census tract. We calculate similar levels of misallocation when the optimal subsidies are allowed to vary by census tract rather than by state.

[6] Shapiro (2021) and Jacobsen et al. (2022) quantify the environmental costs associated with implicit subsidies for less environmentally friendly commodities, such as more carbon-intensive industries and older vehicles. Miller et al. (2019) calculate the optimal geographically-differentiated government subsidies for Medicare Advantage. They also use a policy-function approach to model firm behavior.

Cormick (2018), Sexton et al. (2021), Borenstein and Bushnell (2022), Holland et al. (2022), Lamp and Samano (2023)). Of these papers, ours is closest to Sexton et al. (2021), who estimate the marginal benefits of residential solar panels across the US and calculate the environmental benefits of reallocating solar panels across space.[7] Our primary goal differs from this paper in that we aim to quantify the spatial misallocation and environmental consequences of the current set of subsidies relative to the optimal subsidies. Our contribution is, therefore, to build and estimate a structural model of solar panel demand and electricity production, which we use to quantify the effects of alternative subsidy schemes on the distribution of solar panel installations and calculate the resulting environmental benefits. In doing so, we develop novel, tractable approaches to modeling 1) solar panel demand and 2) power plant production and the associated emissions over space and time.

Our paper is related to several papers which use model-based approaches to quantify the effectiveness of various types of subsidies on inducing solar panel installations (e.g., Burr (2014), De Groote and Verboven (2019), Langer and Lemoine (2022), Feger, Pavanini, and Radulescu (2022)). These papers use rich, dynamic models to study the trade-offs associated with production subsidies versus investment subsidies and the optimal time profile of subsidies. In contrast, our focus is on how subsidies should optimally vary across space and on quantifying the environmental benefits of alternative subsidy schemes.

Finally, this paper is related to the literature estimating the determinants of residential demand for solar panels (e.g., Kwan (2012), Sarzynski, Larrieu, and Shrimali (2012), Kirkpatrick and Bennear (2014), Hughes and Podolefsky (2015), Crago and Chernyakhovskiy (2017), Gillingham and Tsvetanov (2019), Bollinger et al. (2022)).[8] We show that simulations from our structural model are consistent with estimates from this literature. However, our goal is to solve for the optimal subsidies and to simulate the environmental consequences of various subsidy schemes, which requires a structural model of solar panel demand and a model of commercial electricity supply with associated environmental damages. Our paper is the first to combine these two components: an estimated structural model of rooftop solar demand and a model of plant-level electricity production and environmental damages. This aspect of our model allows us to measure the effects of subsidy schemes on installations and

[7] Sexton et al. (2021) and Lamp and Samano (2023) calculate the misallocation of residential solar panels by calculating the benefits of reallocating panels across space, holding constant the total number of panels installed, subject to transmission and local capacity constraints. Here we consider a government that can only influence installations through subsidies. Our exercise, therefore, requires a model of how solar panel installations respond to various subsidy schemes. Further, both these papers assume that the marginal benefits of solar panels across space do not change when panels are reallocated. Our model allows for endogenous changes in power plant production profiles and, therefore, the benefits of solar installations in response to electricity demand and renewable production.

[8] Relatedly, Pretnar and Abajian (2021) calibrate a model of household demand for solar panels and electricity which allows for geographic heterogeneity in preferences.

quantify the benefits of these changes in installations from an environmental perspective.[9]

2 Model

We combine a model of household solar panel demand with a commercial and residential electricity production model. Households are distributed geographically across the United States, and states vary in their electricity prices, installation prices, and the set of subsidies for solar panels they offer. Within states, households vary in their local solar irradiance (sunlight), the amount of space they have for potential solar panels, and their preferences over solar panel installation. Households choose the number of solar panels to install, accounting for electricity and installation prices, solar panel subsidies, and their individual preferences for installing solar panels.

In addition to residential solar panels, commercial power plants also produce electricity. Power plants differ in the extent to which their electricity production leads to environmental damages and their location, which determines how the electricity they produce is distributed across the country. Further, power plants face capacity and non-negativity constraints and vary in the order in which they are dispatched, implying that some power plants will only operate when demand is sufficiently high while others will operate even when demand is low.

2.1 Households

Households, indexed by i, are endowed with income y_i and $\bar{N}_i$ spaces they can potentially use for solar panels. Household i has access to a solar panel technology that can produce a stream of solar energy of $\{A_{it}\}_{t=0}^{T}$ over the lifetime of the panel for each panel they choose to install. In practice, we will think of t as indexing hours and set the lifespan of a solar panel to 25 years.[10] We can think of this solar technology as reflecting the intermittent sunlight profile at a given household's residence, accounting for the depreciation of solar panel efficacy over time.[11] Let j index the state in which the household lives.

Households choose whether or not to install solar panels, $m_i \in \{0, 1\}$, the number of panels conditional on installation, $N_i \in \left(0, \bar{N}_i\right]$, and how much electricity to consume each period. Specifically, households choose a sequence of electricity usage $\{e_{it}\}_{t=0}^{T}$, where e_{it} gives

[9] More broadly, our paper is related to the literature on the aggregate effects of spatial misallocation of production factors (e.g., Albouy (2009), Brandt, Tombe, and Zhu (2013), Herkenhoff, Ohanian, and Prescott (2018), Fajgelbaum et al. (2019), Hsieh and Moretti (2019), Babalievsky et al. (2021)). In this paper, we focus on quantifying the environmental costs of this misallocation rather than the effects on aggregate output.

[10] This is a standard value of the average useful life of solar panels (see e.g., Xu et al. (2018), Chowdhury et al. (2020), or Sodhi et al. (2022)).

[11] For simplicity and because of data limitations, we assume that the solar technology is constant across all of a given household's potential solar panel space.

household i's energy consumption in period t. We assume the household pays a constant price of p_j for all electricity purchased.[12] Let r denote the real interest rate and let $e_i = \sum_{t=0}^{T} \frac{e_{it}}{(1+r)^t}$ denote the discounted sum of energy consumed, such that $p_j e_i$ gives the present discounted cost of electricity consumed.

If a household chooses to install solar panels, they pay the cost of installation of $p_j^{\text{Ins}}(N_i)$, which is a function of N_i, the number of panels they choose to install.[13] The installation cost function $p_j^{\text{Ins}}(\cdot)$ can vary nonlinearly in N_i and is allowed to vary by state j.[14] Households can use electricity generated by solar panels to power their home or can sell it back to the grid. Assume, for now, that households can sell back to the grid at the price of electricity purchased, p_j.[15] Letting $A_i = \sum_{t=0}^{T} \frac{A_{it}}{(1+r)^t}$ denote the discounted sum of electricity production, we can write the present discounted value of energy produced by each solar panel for household i as $p_j A_i$.

Households receive subsidies for solar installations. We allow for three types of solar panel subsidies that capture the majority of state and federal subsidies in the US. First, households can receive a cost-based subsidy s_j^{Cost}, which pays a percentage of the solar installation cost, similar to the federal investment tax credit.[16] Second, households can receive a production-based subsidy of s_j^{kWh} for each kWh of electricity produced by their solar panels, similar to solar renewable energy certificates. Finally, we allow for a per-panel subsidy s_j^{Unit}, such as subsidies that pay per kilowatt of solar capacity installed.

We can thus write the household's budget constraint as

$$c_i + \underbrace{p_j\left(e_i - m_i N_i A_i\right)}_{\text{Net cost of electricity}} + \underbrace{m_i\left(1 - s_j^{\text{Cost}}\right) p_j^{\text{Ins}}\left(N_i\right)}_{\text{Net cost of installation}} = y_i + \underbrace{m_i N_i A_i s_j^{\text{kWh}}}_{\text{kWh Subsidy}} + \underbrace{m_i N_i s_j^{\text{Unit}}}_{\text{Unit Subsidy}} \qquad (1)$$

where c_i is consumption of the numeraire good.[17]

[12] We assume that electricity prices are constant over time. While electricity prices change over time, there is evidence that consumers do not correctly forecast the extent to which prices change over time and expect future prices to be similar to current prices (Hughes and Podolefsky, 2015; Anderson, Kellogg, and Sallee, 2013). Further, we assume these electricity prices are fixed across counterfactuals and, therefore, abstract from the general equilibrium effects of subsidies on electricity prices. Utility companies act as regulated monopolies, which limits their ability to change prices in response to demand-side changes.

[13] We assume a partial equilibrium setting where these installation costs are given exogenously and do not change in response to changes in subsidies.

[14] We assume a nonlinear pricing function to allow for the possibility that there is a fixed cost associated with installing a positive number of panels.

[15] This is consistent with households in states with net metering. In 2017, 39 states mandated net metering policies. Idaho did not have a state net-metering policy, but each of the state's three investor-owned utilities had a net-metering policy. Five other states in our sample have distributed generation rules other than net metering. We discuss how we model households without net metering in Appendix B.1.

[16] In estimation, we will also consider sales tax exemptions and property tax exemptions as cost subsidies.

[17] We can think of c_i and y_i as the present values of consumption and income over time, respectively.

Households have the following quasilinear utility function

$$c_i + \nu_i \left(\{e_{it}\}_{t=0}^{T} \right) + m_i \gamma_i \left(N_i \right),$$

where $\gamma_i(N_i)$ is a strictly concave function which gives the nonpecuniary benefit of adding N_i solar panels for household, and $\nu_i\left(\{e_{it}\}_{t=0}^{T}\right)$ is a function which gives the lifetime utility of electricity usage. The function $\gamma_i(\cdot)$ captures inconvenience costs and any other individual preferences for installing solar panels.[18]

Note that the choice of electricity consumption does not depend on the household's choice to install panels.[19] Thus, we can think of household optimization as a two-step process. First, the household chooses electricity use, $\{e_{it}^{\star}\}_{t=0}^{T}$, then decides whether to install solar panels and the number of panels conditional on installation. In this second stage, we can rewrite the household's optimization problem as a choice of N_i and a discrete choice of m_i:

$$V_i = \max_{N_i, m_i \in \{0,1\}} m_i \left[\mu_{ij} \left(N_i \right) + \gamma_i \left(N_i \right) \right]. \tag{2}$$

where

$$\mu_{ij}\left(N_i\right) = \underbrace{N_i A_i \left(p_j + s_j^{\text{kWh}}\right)}_{\text{Total electricity value}} - \underbrace{\left(1 - s_j^{\text{Cost}}\right) p_j^{\text{Ins}}\left(N_i\right)}_{\text{Net installation cost}} + \underbrace{N_i s_j^{\text{Unit}}}_{\text{Unit subsidy}} \tag{3}$$

denotes household i's net monetary benefit of installing solar panels.[20] Let $m_i^{\star}$ denote the household's optimal installation choice and let $N_i^{\star}$ denote the optimal number of panels conditional on installation.

From equations (2) and (3), we can see that different types of subsidies will differ in the distribution of households they induce to install panels. Households in sunny areas (high A_i) are more likely to respond to the production subsidy s_j^{kWh}, while households in areas with high installation costs are more likely to respond to the cost subsidy s_j^{Cost}, for example.[21] Further, changes in subsidies affect installations via both intensive and extensive margin adjustments: increases in subsidies can increase the number of panels installed for households who are inframarginal with respect to installation and can also induce marginal households who do

[18] We do not model the role of peer-influences in solar panel installations, as documented by Bollinger et al. (2022).

[19] This results from the assumptions that 1) utility is quasilinear and 2) electricity can be bought and sold at the same price.

[20] Note that we have dropped a constant representing the household's utility from electricity use and costs but does not affect the decision to install solar panels.

[21] De Groote and Verboven (2019) find that households heavily discount future benefits associated with solar installations. This result implies that cost-based subsidies may be more effective at inducing installations than production-based subsidies because households receive cost-based subsidies sooner. It would be straightforward to examine our results' robustness to alternative household discount rate values.

not install initially to install a positive number of panels. As we show in Section 2.3, the planner chooses the optimal set of subsidies accounting for these intensive and extensive margin adjustments and for the fact that different households are marginal with respect to each type of subsidy.

2.2 Electricity Production

2.2.1 Background

Before proceeding to the model, we give a brief overview of commercial electricity production in the US. The electricity sector in the US is highly regulated and does not operate like a traditional market. Each of the around 10,000 commercial power plants in the US is overseen by a balancing authority, an entity tasked with matching electricity supply and demand by managing production from individual plants and trading with other balancing authorities. Transmission of electricity between balancing authorities disproportionately occurs within larger regions called NERC *regions*, each constituting a relatively closed market of balancing authorities. Transmission across regions does occur, but this inter-regional transmission occurs almost exclusively within *interconnections*, a geographic unit larger than a region. There are three interconnections in the US: Eastern, Western, and Texas.

We can divide power plants into those that are *dispatchable* and those that are *nondispatchable*. Nondispatchable power sources are those whose output cannot be easily controlled in response to fluctuations in electricity demand and generally produce when available, such as wind and solar.[22] These energy sources are generally intermittent, meaning their productive capacity fluctuates over time in response to environmental factors, e.g., sunlight and wind. Nondispatchable power plants generally do not produce significant pollutants or greenhouse gases.

On the other hand, balancing authorities can control production by dispatchable power plants to satisfy electricity demand. The production profile of a given dispatchable plant is determined by its position in the dispatch curve—the order at which balancing authorities dispatch power plants to satisfy different electricity demand levels.[23] This implies that a power plant's production is not simply proportional to demand—some power plants operate continuously throughout the day while others only operate at peak levels of demand. As such, the set of marginal power plants, and therefore the marginal benefits to residential

[22] Occasionally, dispatch authorities curtail production by solar and wind producers when total production by nondispatchable and baseload units exceeds supply.

[23] Power plants' variable cost of production generally determine the dispatch curve. Power plants with the lowest variable costs (often nuclear and hydroelectric) typically satisfy low demand. Meanwhile, plants with higher variable costs (such as gas-fired plants) begin operating only when electricity demand is sufficiently high.

solar installation, vary geographically and within location as a function of demand that must be satisfied by dispatchable plants.[24]

2.2.2 Model: Electricity Production

Within the model, three sources supply electricity: 1) residential solar, 2) nondispatchable plants, and 3) dispatchable plants.[25] Nondispatchable units are assumed to operate at full capacity conditional on environmental conditions (e.g., sun and wind). Therefore, the production by these power plants is independent of demand and production by other plants.[26] Alternatively, the production by dispatchable units depends on excess demand remaining after production by residential solar and nondispatchable plants.[27]

Residential Solar and Nondispatchable Plants Let R index NERC regions.[28] Total residential solar production in region R in time t is the sum of energy produced by residential solar panels, $E_{Rt}^{\text{Solar}} = \int_{i \in I_R} m_i^\star N_i^\star A_{it} h(i) di$, where $h(i)$ is the measure of households of type i and I_R is the set of households who reside in region R. Similarly, total production by nondispatchable plants in region R in time t is given by $E_{Rt}^{\text{NonD}} = \sum_{k \in K_R} y_{kt}^{\text{NonD}}$, where y_{kt}^{NonD} denotes electricity production by nondispatchable power plant k in time t, and K_R is the set of nondispatchable plants in region R. y_{kt}^{NonD} is allowed to vary fully by power plant k and time t, reflecting differences in environmental factors across plants and over time, and we assume it is independent of demand and production from other power plants.

Dispatchable Plants To capture the centralized manner by which balancing authorities dispatch power plants to satisfy electricity demand, we model dispatchable plants' behavior via policy functions that map excess demand to plant-level production. Let Load_{Rt} denote

[24] See Holland and Mansur (2008) for an investigation of how the environmental costs of electricity production vary as a function of demand.

[25] For most of our analysis, we assume the distribution of power plants and the characteristics of the grid are exogenous. In reality, a large change in residential solar production may lead to the entry and exit of generators and changes in the organization of the electricity grid. In Section 7.4, we analyze the robustness of our results to alternative assumptions about the distribution of power plants. See Holland, Mansur, and Yates (2022) for a model which includes endogenous entry and exit of generators and storage capacity. See Arkolakis and Walsh (2022) for a model with endogenous grid formation.

[26] The exception is when total demand is lower than the amount produced by these nondispatchable generators. In this case, we assume production is curtailed such that supply does not exceed demand.

[27] This assumption is similar to an assumption made by Callaway, Fowlie, and McCormick (2018), who assume that only fossil-fuel power plant production is affected by changes in renewable production.

[28] We will assume 7 NERC regions in our quantitative analysis. Officially, the North American Electric Reliability Corporation (NERC) divides the US into 6 regional entities. Following Holland et al. (2016), we separate California from the WECC region, leaving us with 7 regions. We discuss how we define the regions in Appendix A.7.

the total electricity demand in region R in time t,[29] and let $\text{ELoad}_{Rt} = \text{Load}_{Rt} - E^{\text{NonD}}_{Rt} - E^{\text{Solar}}_{Rt}$ give the electricity demand in region R that is not satisfied by residential solar and nondispatchable plants. We write production by dispatchable plants as a reduced-form function of excess demand across regions, subject to non-negativity and capacity constraints. Letting y^{Disp}_{kt} denote production by dispatchable plant k in time t, we specify

$$y^{\text{Disp}}_{kt} = \begin{cases} 0 & \text{if } f_k\left(\text{ELoad}_t, \varepsilon_{kt}\right) \leq 0 \\ f_k\left(\text{ELoad}_t, \varepsilon_{kt}\right) & \text{if } 0 < f_k\left(\text{ELoad}_t, \varepsilon_{kt}\right) < \bar{y}_k \\ \bar{y}_k & \text{if } f_k\left(\text{ELoad}_t, \varepsilon_{kt}\right) \geq \bar{y}_k \end{cases}, \tag{4}$$

where $\bar{y}_k$ is power plant k's nameplate capacity, the maximum productive capacity of the plant, ELoad_t is the vector of excess loads in each region at time t, and $f_k\left(\text{ELoad}_t, \varepsilon_{kt}\right)$ is a plant-specific function of excess load across regions and a cost shifter ε_{kt}.[30] We allow $f_k(\cdot)$ to depend not only on excess load in the region in which the power plant is located but potentially to depend on excess load across other regions as well. This dependence reflects that electricity can be transmitted across regions in response to excess demand. Note that y^{Disp}_{kt} will be non-linear in ELoad_{Rt} for any given region, given that the plant faces non-negativity and capacity constraints, and because the latent function, $f_k(\cdot)$, is not constrained to be linear in excess load.

2.2.3 Damages

Let $d_k\left(y^{\text{Disp}}_{kt}\right)$ be a function that maps dispatchable power plant k's electricity production in time t to the total environmental damages associated with the plant's emissions of greenhouse gases and air pollutants. Let $D_t\left(\text{ELoad}_t\right) = \sum_k d_{kt}\left(y^{\text{Disp}}_{kt}\right)$ denote total damages from all power plants in time t and let $D\left(\text{ELoad}\right) = \sum_{t=0}^{T} \frac{D_t(\text{ELoad}_t)}{(1+r)^t}$ denote the net present value of all damages over time, where ELoad gives the excess load across all region and time periods.

The external benefit of a marginal solar panel installed by household i equals the damages offset over the panel's lifetime. We write this as

$$\Delta D_i\left(\text{ELoad}\right) \equiv \left|\frac{\partial D\left(\text{ELoad}\right)}{\partial N_i}\right| = \sum_{t=0}^{T} \frac{A_{it}}{(1+r)^t} \left|\frac{\partial D_t(\cdot)}{\partial E^{\text{Solar}}_{Rt}}\right|,$$

[29] This is equal to the sum of household electricity demand plus commercial electricity demand, which we treat as exogenous.

[30] We assume that the plant's policy function depends only on the current excess demand levels. Hypothetically, production could also depend on previous electricity demand and production if, for example, the grid can store significant amounts of electricity over time. We can accommodate this extension by allowing the function $f_k(\cdot)$ to depend on lagged values of excess demand. It would also be straightforward to allow the policy functions to depend on fuel prices or on lagged production levels of the individual plant.

the present discounted sum of the product of A_{it}, the electricity produced by the panel in any given period, and the absolute value of $\frac{\partial D_t(\cdot)}{\partial E_{Rt}^{\text{Solar}}}$, the marginal damages associated with nondispatchable plant production.

2.3 Government's Problem and Optimal Subsidies

The government chooses subsidies to maximize the sum of total utility minus total environmental damages subject to an externally set budget constraint.[31] To ease up on notation, let $s_{ij} = s_j^{\text{Unit}} N_i^\star + s_j^{\text{kWh}} A_i N_i^\star + s_j^{\text{Cost}} p_j^{\text{Ins}}(N_i^\star)$ denote the total subsidy paid to household i conditional on installation. Further, let $\frac{\partial N_i^\star}{\partial s_j^\theta}$ give the derivative of solar panels installed by household i with respect to a given subsidy type $\theta \in \{\text{kWh}, \text{Unit}, \text{Cost}\}$, and let $\overrightarrow{m_i}^\theta$ indicate the household i is on the margin of installing a positive number of panels with respect to a θ subsidy, meaning the household does not install given the current subsidies but would install in response to a small increase in the given subsidy. Finally, let $M_j = \int_{i \in I_j} m_i h(i) di$ denote the total number of households who install solar panels in state j, where I_j is the set of households in state j and again where $h(i)$ is the measure of households type i.

The government maximizes the sum of utility less environmental damages, which we write as

$$\underbrace{\int_i V_i h(i) di}_{\text{Utility}} - \underbrace{D\left(\text{ELoad}\right)}_{\text{Damages}}. \tag{5}$$

The government faces the constraint that the sum of subsidies cannot exceed an externally set constraint

$$\underbrace{\sum_j \int_{i \in I_j} s_{ij} m_i^\star h(i) di}_{\text{Government Cost}} \leq G, \tag{6}$$

where G is the maximum amount the government can spend on subsidies.[32] We can refor-

[31] We are assuming that these are the only policy instruments the government can access. The government is restricted to not price the externality directly, as in Pigou (1920). Changes in subsidies could also change firm profits. We assume that the government does not value profits of utility companies or solar panel installation companies. In reality, utility companies operate as regulated monopolies, where profits are directly limited. Profits of solar panel installation firms not entering the government's objective is also consistent with a model in which the price of installation is always equal to the marginal cost of an installation. We analyze a planner who minimizes environmental damages in Section 6.2. In Appendix B.6, we consider a planner who maximizes the sum of utility subject to a net-cost budget constraint, where environmental damages count as a fiscal cost.

[32] Our framework does not account for distributional effects because utility is quasilinear and Pareto weights are equal across households, which implies equal marginal social welfare weights across households. This setup also implies that household utility is measured in dollar equivalents and, therefore, can be compared directly to environmental damages. See Section 7.5 for a discussion of distributional effects.

mulate the government's objective function as the Lagrangian

$$W = \underbrace{\int_i V_i h(i) di}_{\text{Utility}} - \underbrace{D\left(\text{ELoad}\right)}_{\text{Damages}} - \lambda \left(\underbrace{\sum_j \int_{i \in I_j} s_{ij} m_i^{\star} h(i) di}_{\text{Government Cost}} - G \right), \tag{7}$$

where λ is equal to the marginal value of public funds. In practice, we will set G to the present discounted value of the national cost of solar subsidies, given the current system of subsidies.

The optimal system of subsidies must satisfy $\frac{\partial W}{\partial s_j^{\theta}} = 0$ for each type of subsidy in each state, which implies

$$\underbrace{\frac{\partial M_j}{\partial s_j^{\theta}} \times \left(\overrightarrow{\Delta D}_j^{\theta,\text{ext}} - \lambda \overrightarrow{s}_j^{\theta,\text{ext}} \right)}_{\text{Extensive Margin}} + \underbrace{\left.\frac{\partial N_j}{\partial s_j^{\theta}}\right|_{M_j^{\text{st}}} \times \left(\overrightarrow{\Delta D}_j^{\theta,\text{int}} - \lambda \overrightarrow{\frac{\partial s}{\partial N}}_j^{\theta,\text{int}} \right)}_{\text{Intensive Margin}} + \underbrace{(1-\lambda) M_j \overline{\frac{\partial s_{ij}}{\partial s_j^{\theta}}}}_{\text{Mechanical Effect}} = 0. \tag{8}$$

We provide a derivation for equation (8) in Appendix B.3. The first term ("Extensive Margin") captures the welfare effects associated with households on the margin of installation: the households who currently do not install any solar panels but would install solar panels in response to a slight increase in a given subsidy s_j^{θ}. The term

$$\frac{\partial M_j}{\partial s_j^{\theta}} = \int_{i \in I_j} \overrightarrow{m}_i^{\theta} h(i) di$$

gives the number of households on the margin of installing solar panels with respect to a given subsidy type s_j^{θ}. These marginal installations lead to a societal benefit by reducing environmental damages. The average damages offset across marginal installer households is denoted as $\overrightarrow{\Delta D}_j^{\theta,\text{ext}}$ and is formally given by

$$\overrightarrow{\Delta D}_j^{\theta,\text{ext}} = \frac{\int_{i \in I_j} \Delta D_i \left(\text{ELoad}^{\text{SB}}\right) N_i^{\star} \overrightarrow{m}_i^{\theta} h(i) di}{\int_{i \in I_j} \overrightarrow{m}_i^{\theta} h(i) di},$$

where ELoad^{SB} is the excess load under the optimal system of subsidies. These marginal installations also receive subsidies and thus are associated with a fiscal cost. We denote the average cost associated with a marginal installation household as $\overrightarrow{s}_j^{\theta,\text{ext}}$, formally written as

$$\overrightarrow{s}_j^{\theta,\text{ext}} = \frac{\int_{i \in I_j} s_{ij} \overrightarrow{m}_i^{\theta} h(i) di}{\int_{i \in I_j} \overrightarrow{m}_i^{\theta} h(i) di}.$$

The second term of equation (8) ("Intensive Margin") captures the welfare effects associated with intensive margin adjustment: increases in the number of panels purchased for households who already choose to install a positive number of panels. The term $\frac{\partial N_j^{\text{st}}}{\partial s_j^{\theta}}\big|_{M_j^{\text{st}}}$ gives the total increase in panels associated with an increase in a given subsidy, holding the set of households who install solar panels constant, which we write as

$$\left.\frac{\partial N_j^{\text{st}}}{\partial s_j^{\theta}}\right|_{M_j^{\text{st}}} = \int_{i \in I_j} m_i^{\star} \frac{\partial N_i^{\star}}{\partial s_j^{\theta}} h(i) di.$$

The terms $\overrightarrow{\Delta D}_j^{\theta,\text{int}}$ and $\overrightarrow{\frac{\partial s}{\partial N}}_j^{\theta,\text{int}}$ give the average damages offset and the average fiscal cost, respectively, associated with these marginal panels.[33] Taken together, these first two terms show that the government will optimally increase subsidies which induce a greater number of installations and more panels per installation from households associated with significant environmental benefits and for whom current subsidies are low.

The final term ("Mechanical Effect") captures the effects of increasing subsidies for the households who already choose to install solar panels and thus receive a larger subsidy from the government. The total size of this transfer is the total number of panels installed in state j, M_j, multiplied by the average increase in subsidy for households who have installations, $\overline{\frac{\partial s_{ij}}{\partial s_j^{\theta}}}$, holding installations and the number of panels constant.[34] Each dollar transferred to these inframarginal households increases welfare by $(1 - \lambda)$, which reflects the increase in household utility less the decrease in government funds.[35]

However, solving for optimal subsidies requires more structure on the problem. While it may be possible to calculate the environmental benefits of marginal solar panel installations given the current distribution of solar panel installations, to solve for the optimal subsidies, we need to know how marginal benefits change in response to different subsidy schemes.

[33]These are formally given by $\overrightarrow{\Delta D}_j^{\theta,\text{int}} = \frac{\int_{i \in I_j} \Delta D_i(\text{ELoad}^{\text{SB}}) m_i^{\star} \frac{\partial N_i}{\partial s_j^{\theta}} h(i) di}{\int_{i \in I_j} m_i \frac{\partial N_i}{\partial s_j^{\theta}} h(i) di}$ and $\overrightarrow{\frac{\partial s}{\partial N}}_j^{\theta,\text{int}} = \frac{\int_{i \in I_j} \frac{\partial s_{ij}}{\partial N_i} m_i \frac{\partial N_i}{\partial s_j^{\theta}} h(i) di}{\int_{i \in I_j} m_i \frac{\partial N_i}{\partial s_j^{\theta}} h(i) di}$.

[34]Formally this is $\overline{\frac{\partial s_{ij}}{\partial s_j^{\theta}}} = \frac{\int_i m_i^{\star} \frac{\partial s_{ij}}{\partial s_j^{\theta}} h(i) di}{\int_i m_i^{\star} h(i) di}$.

[35]Note that the utility of marginal households does not show up in equation (8) since there is no first-order welfare effect on households for marginal households (i.e., households who choose to install solar panels in response to the increase in subsidies). This result comes from the envelope theorem. See also Colas, Findeisen, and Sachs (2021) for a discussion of the roles played by marginal and inframarginal agents in the first-order effects of targeted subsidy increases.

Given that power plants' production plans are nonlinear, the marginal damages evaluated at current installation levels will differ from those at the optimum. Further, the optimal subsidies characterized by equation (8) depend not only on marginal damages, but also on the number of inframarginal households and the number of households on the margin of installation with respect to various types of subsidies. Like the marginal damages, both of these objects are a function of the system of subsidies.

Therefore, our approach is to estimate a fully specified version of our model, and then use that model to quantify the system of optimal subsidies. Further, the structural model allows us to quantify the effects of alternative subsidy schemes on the spatial distribution of rooftop solar and the resulting environmental benefits.

3 Quantitative Model

3.1 Household preferences

Recall that we can summarize the household's choice to install solar panels as follows,

$$V_i = \max_{N_i, m_i \in \{0,1\}} m_i \left[\mu_{ij}\left(N_i\right) + \gamma_i\left(N_i\right)\right],$$

where

$$\mu_{ij}\left(N_i^{\star}\right) = \underbrace{N_i A_i \left(p_j + s_j^{\text{kWh}}\right)}_{\text{Total electricity value}} - \underbrace{\left(1 - s_j^{\text{Cost}}\right) p_j^{\text{Ins}}\left(N_i\right)}_{\text{Net installation cost}} + \underbrace{N_i s_j^{\text{Unit}}}_{\text{Unit subsidy}}.$$

Let ℓ denote the census tract in which a household lives. We assume that the nonpecuniary utility of installation of installing N_i panels, $\gamma_i\left(N_i\right)$, is given by a polynomial term in N_i, a term that captures differences in the benefits of installation across demographic groups, and an idiosyncratic term. Specifically, we parameterize the nonpecuniary value of installation as

$$\gamma_i\left(N_i\right) = \underbrace{\gamma_0 + \gamma_{1N} N_i + \gamma_{2N} N_i^2}_{\text{Polynomial in } N_i} + \underbrace{\gamma_{dem} X_\ell}_{\text{Local Demographics}} + \underbrace{\sigma \epsilon_i}_{\text{Idiosyncratic}}$$

where γ_0, γ_{1N}, and γ_{2N} are parameters, γ_{dem} is a vector of parameters, X_ℓ is a vector of demographic characteristics associated with the tract in which the household lives, and ϵ_i is a logit preference draw with scaling parameter σ.[36] In practice, we specify $\gamma_{dem} X_\ell = \gamma_{Coll} X_\ell^{Coll} + \gamma_{Pol} X_\ell^{Pol}$, where X_ℓ^{Coll} the fraction of individuals in the census tract with a college education and X_ℓ^{Pol} is the fraction of voters in the county who voted democrat in the

[36] X could contain household level covariates. However, we only have installation data at the tract level. We can therefore think of X_ℓ as capturing local attitudes towards installation.

2016 election.[37]

Solving for the optimal number of panels conditional on installation yields

$$N_i^\star = \min\left[\bar{N}_i, -\left(\frac{\frac{\partial \mu_{ij}}{\partial N_i} + \gamma_{1N}}{2\gamma_{2N}}\right)\right]. \tag{9}$$

Loosely speaking, we can see that the ratio $\frac{\gamma_{1N}}{\gamma_{2N}}$ dictates the average size of installations while the parameter γ_{2N} dictates the degree to which $N_i^\star$ varies with subsidies.[38] For example, a smaller value of γ_{2N} in absolute value would imply that households are more responsive to subsidies along the intensive margin.

Given that draws of ϵ_i are from a logit distribution, the probability that a household installs panels is equal to

$$\pi_i = \frac{\exp\left(\frac{\mu_{ij}\left(N_i^\star\right)+\gamma_0+\gamma_{1N}N_i^\star+\gamma_{2N}N_i^{\star 2}+\gamma_{dem}X_\ell}{\sigma}\right)}{1+\exp\left(\frac{\mu_{ij}\left(N_i^\star\right)+\gamma_0+\gamma_{1N}N_i^\star+\gamma_{2N}N_i^{\star 2}+\gamma_{dem}X_\ell}{\sigma}\right)}. \tag{10}$$

The partial elasticity of installation probability with respect to monetary benefits is equal to

$$\frac{\partial \log\left(\pi_i\right)}{\partial \mu_{ij}\left(N_i^\star\right)} = \frac{1}{\sigma}\left(1-\pi_i\right).$$

Therefore, the parameter σ dictates the extent to which increases in subsidies will lead to increases in installations. A smaller value of σ implies that increases in subsidies will lead to larger increases in the number of installations.

3.2 Dispatchable Power Plant Production

Recall from equation (4) that production by dispatchable power plant k in time t is

$$y_{kt}^{\text{Disp}} = \begin{cases} 0 & \text{if } f_k\left(\text{ELoad}_t, \varepsilon_{kt}\right) \le 0 \\ f_k\left(\text{ELoad}_t, \varepsilon_{kt}\right) & \text{if } 0 < f_k\left(\text{ELoad}_t, \varepsilon_{kt}\right) < \bar{y}_k \\ \bar{y}_k & \text{if } f_k\left(\text{ELoad}_t, \varepsilon_{kt}\right) \ge \bar{y}_k. \end{cases}$$

[37] We examine the robustness of our results to alternative specifications of this utility function in Section 7.1. It would be straightforward to allow $\gamma_i(\cdot)$ to depend on additional tract-level covariates or to let γ_0 vary by Census region, division, or state. Note that if we allow γ_0 to vary by state, then the variance of ε_i, and therefore the responsiveness of the household installation decision to monetary incentives, would be identified solely by variation in solar irradiance within states, not by variation in prices and subsidies across states.

[38] As we explain below, we parameterize $p_j^{\text{Ins}}(N_i)$ as a fixed cost plus a constant per-panel cost. This implies that $\frac{\partial \mu_{ij}}{\partial N_i}$ will be constant for a given household for $N_i > 0$.

We assume the latent function $f_k(\text{ELoad}_t, \varepsilon_{kt})$ is a second-degree polynomial of excess load in each NERC region within plant k's interconnection with an additive shifter denoted by ε_{kt}. Formally, letting $\mathbf{R}_k$ denote the set of NERC regions within the interconnection that contains plant k, we specify

$$f_k(\text{ELoad}_t, \varepsilon_{kt}) = \psi_k^0 + \sum_{R \in \mathbf{R}_k} \left(\psi_{Rk}^1 \text{ELoad}_{Rt} + \psi_{Rk}^2 \text{ELoad}_{Rt}^2\right) + \varepsilon_{kt}, \tag{11}$$

where ψ_k^0 is a constant term, ψ_{Rk}^1 is a parameter which dictates how $f_k(\cdot)$ changes in response to excess load in region R, ψ_{Rk}^2 is a parameter which dictates how $f_k(\cdot)$ responds to excess load squared in region R, and ε_{kt} is a normally distributed idiosyncratic term with a mean of 0 and a variance of σ_k^2. Note that all the ψ_k parameters and σ_k^2 are plant-specific. We allow $f_k(\cdot)$ to depend on excess demand in all regions within an interconnection but not on excess demand in other interconnections. This dependence reflects that electricity can be transmitted across regions within interconnections but is rarely transmitted across interconnections.[39] Our assumed functional form implies that y_{kt} is a Tobit function with latent variable f_k that is right censored at $\bar{y}_k$, plant k's nameplate capacity, and left censored at 0.

This specification allows for relatively complex production patterns as a function of excess demand. The parameter ψ_k^0 dictates the values of excess demand over which a plant will produce electricity, allowing for the possibility that some plants will operate when excess demand is low while others will operate when excess demand is sufficiently high. For example, all else equal, plants for which ψ_k^0 takes a large negative value will only have positive electricity production when excess demand is very high, as is true of plants that tend to have a late position in the dispatch curve. Conditional on positive production, the parameters ψ_{Rk}^1 and ψ_{Rk}^2 will dictate the intensity at which the balancing authority dispatches a plant. Finally, this specification allows plants to differ in the extent to which their production is transmitted across regions. Some plants may predominantly transmit power within their own region, while others may transmit large amounts of power to other regions within an interconnection.

It is worth discussing how our specification of power plant production differs from the specifications used in Holland et al. (2016) and Sexton et al. (2021). Those papers estimate plant-level marginal emissions rates with respect to electricity demand that vary across time (for example, by the hour of the day). Their specification allows for flexible estimation of marginal emissions rates, but those estimated rates for each power plant are constant conditional on time.[40] However, since we aim to estimate marginal emissions both under

[39] We constrain $f_k(\cdot)$ such that the function is weakly increasing in excess load for all regions. That is, we set $f_k(\cdot)$ to its value at the inflection point if the function would otherwise be decreasing in excess load.

[40] An exception to this approach is Borenstein and Bushnell (2022), who estimate emissions as a piece-wise linear function in load with separate slopes for each tercile of load.

current conditions and under significant changes to the distribution of residential solar panels, we require a different approach to modeling power plants. Our model allows production to vary flexibility with excess load at the cost of reduced flexibility over time. Thus, marginal emissions are not constant as a function of residential solar production. An additional benefit of our approach is that we identify the model's parameters with excess load, which takes advantage of variation in both demand and production from nondispatchable units. The other models only leverage variation in demand.

3.3 Damages

The final piece of the model is determining damages from electricity production at power plant k, as described by the function $d_k(y_{kt})$. We specify this function in two parts, first mapping electricity generation into emissions and then mapping emissions to damages. Both parts are plant-specific, capturing that damages from electricity production depend on a power plant's technology, location, and stack height. A power plant's technology dictates the extent to which electricity production leads to emissions, while a plant's location and stack height determine the extent to which emissions of local pollutants affect population centers.

Concretely, let $g \in G$ index pollutants, where we assume the set G consists of the pollutants NOx, PM2.5, SO2, and CO2 equivalent (CO2e).[41] We specify emissions of each pollutant as a power-plant-specific linear spline in production with a slope that differs above and below power plant k's median production. Letting y_k^{50} denote the median amount of power plant k's production in the data conditional on positive production, we write power plant k's emissions of pollutant g as

$$\text{Emis}_{gkt}(y_{kt}) = \kappa_{gk}^1 y_{kt} + \kappa_{gk}^2 \left(y_{kt} - y_k^{50}\right)_+ + e_{gkt},$$

where

$$\left(y_{kt} - y_k^{50}\right)_+ = \begin{cases} (y_{kt} - y_k^{50}) & \text{if } (y_{kt} - y_k^{50}) \geq 0 \\ 0 & \text{if } (y_{kt} - y_k^{50}) < 0 \end{cases}.$$

Power plant k's damages in time t are then given by

$$d_{kt}(y_{kt}) = \sum_{g \in G} \delta_{gk} \text{Emis}_{gkt}(y_{kt}),$$

where δ_{gk} gives the marginal damages associated with emissions of g by power plant k,

[41] CO2 equivalent includes emissions of other greenhouse gasses in addition to carbon dioxide, in particular, methane and nitrous oxide. These other GHGs are converted into a common global warming potential equal to that of one ton of carbon dioxide.

accounting for power plant k's location and stack height.

4 Data and Estimation

4.1 Data Sources

In this section, we give an overview of the main data sources we use in our analysis. Additional details on data sources and cleaning can be found in Appendix A.

Solar Panel Installations Our primary source for solar panel installations is the Deepsolar database (Yu et al., 2018), a database of solar panel installation in the contiguous US created by applying a deep-learning model for detecting solar panels on satellite imagery.[42] From Deepsolar, we use tract-level data on the total number of residential solar systems and on the total panel area covered by residential solar panels. Combining these two measurements gives us the average size of solar installations, which we use to infer the average number of panels per installation in each tract.

We supplement these data on solar installations with data from Google Project Sunroof (GPS), another dataset created by applying a machine-learning framework to satellite imagery. This dataset provides the distribution of rooftop sizes that are suitable for solar panel installation in each tract, which we use as the empirical analog of $\bar{N}_i$ within each tract for 56,940 census tracts in the US.[43]

Rooftop Solar Production Next, we need data on $\{A_{it}\}_{t=0}^{T}$, the stream of electricity potentially produced by each panel installed by household i. For this, we combine data on yearly solar production potential from GPS with state-level time profiles of solar production from the National Renewable Energy Laboratory's System Advisor Model (SAM). Specifically, GPS provides measures of yearly kWh that can be produced by panels in a given tract, accounting for local weather conditions and shading. We set a household's yearly solar potential for newly installed panels as the mean household solar potential in the GPS data for the household's tract. We assume solar panel efficacy depreciates by a constant rate of

[42] Deepsolar is the first high-fidelity database of solar panel installations in the United States. Other solar panel databases rely on either self-reported data or surveys (e.g., Open Solar Project) or do not cover the entire contiguous US (e.g., Tracking the Sun). The machine-learning algorithm employed by Deepsolar is highly accurate, achieving a precision of 93% and a recall of 89% in residential areas.

[43] These tracts include 90% of the 33 million square meters of residential solar panels in the Deepsolar database. The GPS data specifically provide the number of buildings in each tract with the potential for various installation size bins. We set $\bar{N}_i$ as the midpoint of the installation size bin for all buildings which fall in a given bin.

0.5% each year.[44]

Next, we need to determine the distribution of solar production over each hour of the panel's lifetime. For this, we utilize SAM, which provides engineering estimates of electricity production with panel specifications and climate as the model inputs (Blair, Dobos, and Gilman, 2013). For each state, we calculate the fraction of yearly solar production produced at any given hour over the year. See Appendix A.4 for details. We multiply this fraction of energy produced each hour by a household's annual solar potential to calculate our measure of A_{it}, hourly electricity production for any hour t over the panel's lifetime.

Subsidies and Prices For subsidies, we rely on data from Sexton et al. (2021), who assemble data from the Database of State Incentives for Renewables & Efficiency (DSIRE) to calculate state and federal subsidies. For the price of electricity, we use the average retail price of electricity as reported by the EIA.[45] We use a value of $r = 2\%$ for the real interest rate.

We estimate installation prices using Tracking the Sun, a project collecting data on solar panel installations by the Lawrence Berkeley National Lab. As Tracking the Sun only covers 25 states, we assume that all states within a given Census region share the same installation pricing function. Specifically, we assume that installation prices take the form $p^{\text{Ins}}_{R(j)}(N_i) = p^{0,\text{Ins}}_{R(j)} + N_i p^{1,\text{Ins}}_{R(j)}$, where $p^{0,\text{Ins}}_{R(j)}$ is a fixed cost and $p^{1,\text{Ins}}_{R(j)}$ is a per-panel cost, and $R(j)$ is the Census region containing state j. We present our estimates of the installation price functions and provide evidence that this linear pricing function is a good approximation of prices in the data in Appendix C.1.

Power Plants Our electricity generation data come from Open Grid Emissions (OGE), an open-source project aimed at creating high-quality electricity emissions data that is publicly available (Miller et al., 2022). These data combine commonly used electricity data sources, namely hourly electricity generation and emissions for generating units from the EPA's Clean Air Markets Division (CAMD), monthly production and emissions for generating units from EIA form 923, and hourly balancing authority by fuel type electricity generation from EIA form 930.

We use their power-systems-level and plant-level data products from 2019. The power-systems-level data gives hourly electricity production for each balancing authority, broken out by fuel category, enabling us to calculate each region's total hourly load. The plant-level data gives hourly electricity production and emissions for just under 10,000 power plants.

[44] Jordan and Kurtz (2013) review the literature on photovoltaic degradation rates and find a median degradation rate of 0.5%.

[45] These data can be downloaded at `https://www.eia.gov/electricity/state/`.

This coverage is the main innovation of the OGE data, as previously hourly emissions and production were only available for sufficiently large fossil-fuel plants included in the EPA's CAMD data. The plants excluded from CAMD data account for nearly 30 percent of NOx emissions, 8 percent of SO2 emissions, and 7 percent of CO2 emissions. We use the 4,625 dispatchable plants with postive, non-constant production in estimation, yielding over 40 million plant-hour observations after the cleaning process described in detail in Appendix A.7.

Damages To calculate damages associated with emissions, we utilize AP3, a state-of-the-art integrated assessment model that translates emissions from locations across the US into physical and economic damages. Specifically, AP3 uses a reduced-complexity air quality model to map emissions of local pollutants to an ambient concentration of air pollutants in each county in the US. The model then translates these ambient concentrations into damages, using estimates of the physical effects of pollution exposure from the literature and considering population distribution and vital statistics across counties.[46] AP3 and its predecessors, APEEP and AP2, have been employed extensively in the environmental economics literature (see, e.g., Muller, Mendelsohn, and Nordhaus (2011), Siler-Evans, Azevedo, and Morgan (2012), Holland et al. (2016), Shapiro and Walker (2020), Holland et al. (2020), Sexton et al. (2021), Cicala et al. (2021), Holland et al. (2021)).

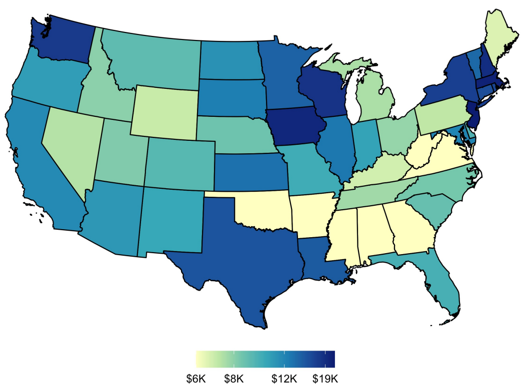
4.2 Descriptive Patterns

Figure 1a shows how the generosity of subsidies varies across states under the current system of subsidies. We measure subsidy generosity as the present discounted value of subsidies an average household in each state would receive if they installed a 15-panel system, roughly the average size of installations in the data.[47] There is considerable variation across states in the generosity of these subsidies. New Jersey delivers nearly 29 thousand dollars in subsidies, compared to seven states providing no additional funding, leading to under 6 thousand dollars in subsidies from the federal government.[48] Figure A4 in Appendix A.6 shows the state-level subsidy generosity separately for each of the three subsidy types. The majority of the value of subsidies comes from cost-based, rather than production-based or unit-based, subsidies.

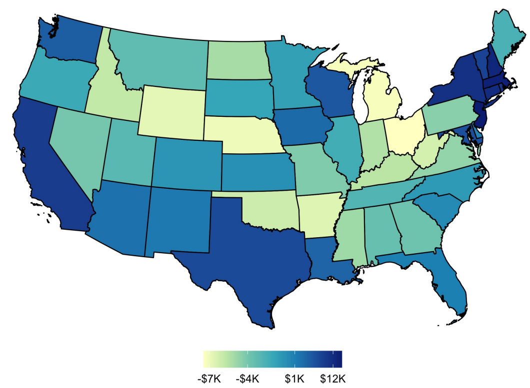
[46]AP3 calculates damages as increased mortality risk from pollution exposure. For the value of mortality risk reduction, we use the EPA's suggested value of \$7.4 million translated into 2014 dollars.

[47]This amount will vary within a given state, as per-kWh subsidies depend on the local level of solar irradiance. We take the population-weighted average across tracts within each state.

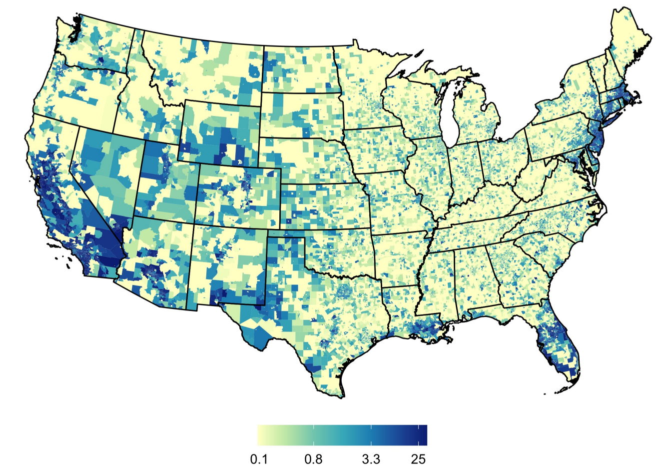
[48]Other states with high subsidies include Massachusetts, Iowa, New Hampshire, Wisconsin, Washington, and New York, with over 18 thousand dollars in expected subsidies. Alabama, Arkansas, Georgia, Mississippi, Oklahoma, Virginia, and West Virginia are the seven states with no state-level subsidies.



(a) Expected subsidy for a 15-panel system



(b) Expected monetary benefit of a 15-panel system



(c) Installed solar systems per 1000 individuals from Deepsolar

Figure 1: Expected subsidies, monetary benefit, and installations. Colors are scaled by the percentile of their respective value. See text for details.

In addition to subsidies, the monetary incentives to install panels vary geographically because of spatial differences in prices and sunlight.[49] Figure 1b shows the monetary benefits associated with solar panel installations. Specifically, for every household within the model, we calculate the net present value of monetary benefits of installation, $\mu_{ij}\left(N_i^\star\right)$, evaluated at $N_i^\star = 15$. We then take the average monetary benefit over all households within a state. This total monetary benefit therefore measures the net present value of installing solar panels in a given state for the average household, taking into account local differences in solar irradiance, electricity and installation prices, and the set of local subsidies. The states with the highest monetary benefits are located in the Northeast, a region with high electricity prices and subsidies. Additionally, California has a high monetary value of installation, combining high electricity prices with high levels of solar irradiance. Meanwhile, several states in the Midwest and Mountain West have negative values, driven by lower subsidies, electricity prices, and solar radiation.

Figure 1c shows installations per capita at the census tract level. We can see that installations are generally higher in areas with larger monetary benefits, such as most of the Northeast and California.[50] Meanwhile, households in the Midwest, where there are relatively low subsidies, less sunlight, and low electricity prices, install few solar panels.

4.3 Estimation

Households We estimate the household installation component of the model via the generalized method of moments. In essence, we choose the six parameters $\sigma, \gamma_0, \gamma_{Coll}$, γ_{Pol}, γ_{1N} and γ_{2N} such that the distribution of installations and size of installations predicted by the model are close to those in the data. Specifically, we target the log installations per household in each census tract and the average number of panels per array in each census tract. Therefore, we have one moment per tract which describes the number of installations, and one moment per tract which describes the number of panels per array.

The six parameters of interest are well identified. Variation in the monetary benefit of installation and demographics across tracts jointly identifies σ, γ_0, γ_{Coll}, and γ_{Pol}. To see this concretely, note from equation (10) that we can rewrite a household's log installation probability as a function of structural parameters:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \frac{\gamma_0}{\sigma} + \frac{1}{\sigma}\mu_{ij}\left(N_i^\star\right) + \frac{\gamma_{Coll}}{\sigma}X_i^{Coll} + \frac{\gamma_{Pol}}{\sigma}X_i^{Pol}. \tag{12}$$

Here it is clear that σ is identified by the relationship between the installation rates and the

[49] We show a map of state-level electricity prices in Appendix A.5.

[50] We show a scatterplot between state-level installations and benefits in Appendix C.2.

monetary benefits of installation, holding demographic characteristics constant.[51] Differences in prices, subsidies, and solar irradiance across tracts drive the variation in monetary benefits. Conditional on monetary benefits, variation in college percentage and democrat percentage identify γ_{Coll} and γ_{Pol} respectively, and the overall level of installations identifies γ_0. Finally, the average number of panels in each array and how the size of arrays varies across cities identify γ_{1N} and γ_{2N}.

Dispatchable Power Plants We estimate the power-plant-specific policy functions described by equations (4) and (11) via maximum likelihood. We provide the likelihood function and additional details in Appendix B.2. Variation over time in both electricity demand and production by nondispatchable plants creates variation in excess loads across regions that identifies the parameters of the plant-specific policy functions.

Damages We estimate damages by combining power-plant level emissions data from EPA's Clean Air Markets Division with estimates of marginal damages from AP3. Recall that power plant k's emissions of pollutant g are given by

$$\text{Emis}_{gkt} = \kappa^1_{gk} y_{kt} + \kappa^2_{gk} \left(y_{kt} - y^{50}_k \right)_+ + e_{gkt}.$$

We estimate these splines via ordinary least squares using power-plant level emissions data from OGE.

To translate these emissions into damages, we need an estimate of δ_{gk}, the marginal damages associated with emissions of pollutant g by power plant k. The AP3 model calculates the marginal damages associated with local pollutants emitted from every county in the United States for varying stack heights. We, therefore, calculate δ_{gk} by matching power plants to their corresponding county and stack height in the AP3 model. We assume a social cost of carbon of 51 dollars per ton of CO2, in line with the official value currently used by

[51] One might be concerned that households with stronger preferences for solar installation tend to live in states with more generous subsidies, and therefore, the correlation between installations and subsidies reflects differences in household characteristics across states rather than the causal effects of subsidies on installations. We do not think this is a serious issue for two reasons. First, as we show in Section 5.1.2, our estimates are consistent with those in Crago and Chernyakhovskiy (2017) and Hughes and Podolefsky (2015), two design-based studies which estimate the effects of subsidies on installation. Crago and Chernyakhovskiy (2017) directly control for pro-environmental preferences while Hughes and Podolefsky (2015) utilize differences in subsidies across a geographic boundary that divided utility companies. Second, in Appendix D.1, we reestimate the model and recalculate the optimal subsidies in models in which we include different sets of tract-level characteristics in the household utility function. The results are not sensitive to which characteristics are included, suggesting that, at least, selection on *observables* does not play an important role in explaining differences in tract-level installation rates. We also estimate the relationship between tract-level installations and monetary benefits in Appendix C.2. We find that the estimates are not sensitive to the inclusion of demographic characteristics and Census region or division fixed effects.

		Estimate	Standard Error
Dispersion of Idiosyncratic Utility	σ	8.55	0.08
Percent College	γ_{Coll}	5.71	0.36
Percent Democrat	γ_{Pol}	10.44	0.50
Constant	γ_0	-1357.03	411.27
Number of Panels	γ_{1N}	177.70	55.65
Number of Panels Squared	γ_{2N}	-6.00	1.88

Table 1: Parameter estimates for household utility function. Standard errors calculated via bootstrapping.

the U.S. government.[52]

5 Estimation Results and Model Fit

5.1 Households

5.1.1 Parameter Estimates

Table 1 displays the estimates of parameters governing the household utility function. The nonpecuniary value of installations is increasing in average local education and in the fraction of the population that voted democrat in the 2016 election. The final two parameter estimates, which dictate utility as a function of installation size, imply that the optimal size of an installation is increasing in monetary benefits, but only marginally so: a \$1000 increase in the monetary benefit associated with installing an additional panel leads to only a $\left|\frac{1}{2\times(-6.00)}\right| \approx 0.08$ increase in the optimal number of panels.[53]

To get a better sense of what the parameter estimates imply for installation probabilities, recall that the partial elasticity of installation probability with respect to monetary benefits is

[52] There is disagreement in the literature about the social cost of CO2. Rennert et al. (2022), for example, argue for a social cost of carbon of 185 dollars per ton of CO2. Calculating the optimal subsidies under alternative values of the social cost of carbon would be straightforward. We do not account for the environmental damages associated with producing and disposing of solar panels. These costs are small relative to the environmental benefits of power produced by a solar panel. Heath and Mann (2012) harmonize the results from 400 studies on the life cycle environmental costs of solar panels. They find that solar panels lead to 40g of CO2 equivalent emissions per kWh of electricity produced throughout their life cycle. These lifecycle emissions imply an environmental cost of roughly \$14 for the average panel in our model compared to an environmental benefit of over \$300. Further, the environmental cost associated with production is unlikely to vary depending on the installation location of the solar panel. In contrast, the environmental benefits we account for here can vary significantly depending on the panel's installation location. It would be straightforward to account for these damages associated with the production and disposal of panels in our model.

[53] Recall from (9) that the optimal number of panels is given by $N_i^\star = \min\left[\bar{N}_i, -\left(\frac{\frac{\partial \mu_{ij}}{\partial N_i}+\gamma_{1N}}{2\gamma_{2N}}\right)\right]$. We estimate $\gamma_{2N} = -6.00$ and increasing the per-panel subsidy by \$1000 increases $\frac{\partial \mu_{ij}}{\partial N_i}$ by 1.

approximately equal to $\frac{1}{\sigma}$.[54] Given that we measure monetary values in thousands of dollars, our estimate of $\sigma = 8.55$ implies that a thousand dollar increase in the net present value of subsidies leads to approximately a $\frac{1}{8.55} \approx 11.7$ percent increase in the number of installations.

5.1.2 Comparison to Existing Estimates

Hughes and Podolefsky (2015) estimate the effects of subsidies on solar panel installations by examining the introduction of a solar rebate in California. In their preferred estimate, they find that a \$470 increase in total rebate leads to a 10% increase in installations. From our estimates above, we can see that a \$470 increase in subsidies would lead to approximately a $.47 \times \frac{1}{\sigma} \approx 5.5$ percent increase in the number of installations. Thus, our result is smaller than the estimate in Hughes and Podolefsky (2015) but of a similar magnitude.

Crago and Chernyakhovskiy (2017) analyze the effects of policy incentives on residential solar panel installations using county-level panel data from 12 states in the US Northeast. They find that increasing rebates by \$1 per watt increases solar panel installations by 47%. We replicate this experiment using our structural model and find that increasing rebates by \$1 per watt in the same 12 states increases installations by 49%, consistent with the estimates in Crago and Chernyakhovskiy (2017).[55]

5.1.3 Model Fit (Installations)

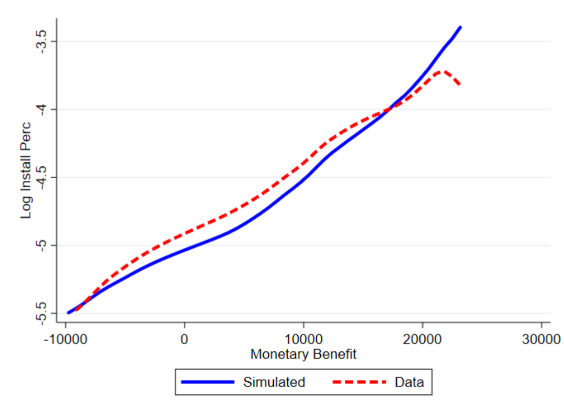
Figure 2 assesses model fit with regard to solar installations. Figure 2a shows the relationship between tract-level log installations per household and the monetary benefits of installation in the data and simulation. We calculate the lifetime monetary benefits of installation as the net present value of installing a 15-panel array in each census tract. We can see that in both the data and simulations, installations are strongly increasing in monetary incentives.[56] Subfigures 2b and 2c show the relationship between installations and the percentage of households with a college education, and the percentage of households who voted democrat in the 2016 election. In both simulations and the data, installations are increasing in educational attainment and democrat percentage. The fit is quite good in both dimensions.

Subfigure 2d shows the relationship between state-level log installation rates in the data and predicted by the model. Each circle represents a state, and each circle's size is proportional to the number of households. The X-axis gives the log installations per household in

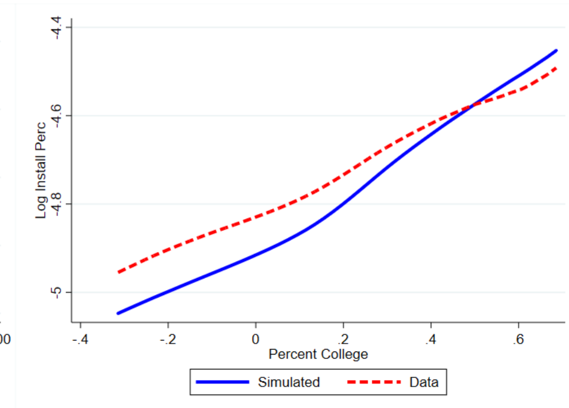
[54]Differentiating (10) yields $\frac{\partial \log(\pi_i)}{\partial \mu_{ij}(N_i^\star)} = \frac{1}{\sigma}(1 - \pi_i)$. The average value of π_i in our dataset is about 0.02.

[55]Specfically, we increase the per-unit subsidy, s_j^{Unit}, in these same 12 states. We convert the per-watt subsidy into a per-unit subsidy by assuming 250 watts per panel.

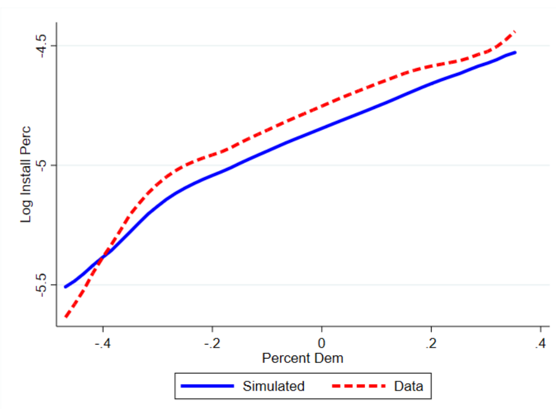
[56]The slight decrease in installation rates for the highest monetary values in the data reflects that Massachusetts and New Jersey have very generous subsidies and high electricity prices, but installation rates are lower than in states such as California and Arizona, which have less generous subsidies.



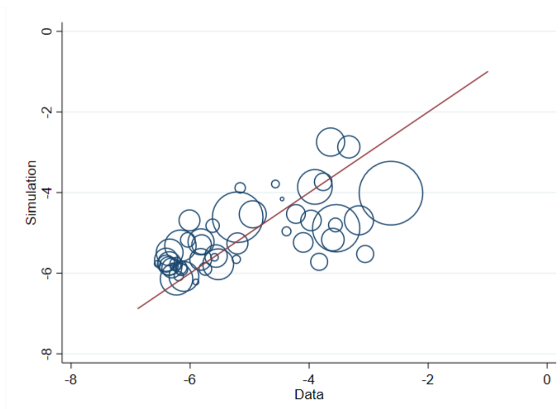
(a) Financial Benefit



(b) Education



(c) Percent Democrat



(d) State-level fit

Figure 2: Panels (a), (b), and (c) show local nonparametric fit of tract-level log installation per household in the data and simulations on various tract-level measures. All panels use an Epanechnikov kernel with bandwidth equal to the standard deviation of the variable on the X-axis. The three panels give the relationships between installations and (a) the total monetary benefit of installing 15 solar panels, (b) the percent of households with a college degree, and (c) the percent of households who are democrats. Panel (d) gives state-level fit. The X-axis gives state-level log installation per household in the data and the Y-axis gives state-level log installations per household in the simulated model. Each circle represents a state and the size of the circle is proportional to the number of households in the state.

the data, while the Y-axis gives the simulated installations per household. In general, the fit is quite good, especially considering we parameterized the model of household panel demand quite sparsely.

5.1.4 Why Do Installations Vary Across States?

In the model, solar panel installation rates may differ across states for five main reasons: 1) subsidies, 2) electricity prices, 3) installation prices, 4) sunlight, and 5) household demographics. In Appendix C.4 we sequentially equalize each of these five factors across states and re-simulate the model. The results from this model-based decomposition show that differences in subsidies and electricity prices across states are the most important drivers of differences in installation rates: equalizing subsidies and electricity prices reduces the standard deviation in state installation rates by 75 percent.

5.2 Power Plants

We now evaluate the performance of the power plant portion of the model. We include additional model fit results for power plants in Appendix C.5.

Figure 3 shows predicted electricity production of dispatchable plants against actual electricity production. Each dot represents an hour of aggregate production by dispatchable plants for each interconnection in the data (X-axis) and predicted by the model (Y-axis). The model fits well in all three interconnections, producing R-squared values of 0.99, 0.96, and 0.96 in the East, West, and Texas, respectively.

We assess fit over hours and seasons in Figure 4. Each panel shows predicted and actual production of dispatchable plants for the average day for each of the three interconnections across four seasons. This figure shows that the model matches daily peaks and troughs of production in response to changes in demand and differences in intraday timing of those peaks and troughs between seasons. A region-level breakdown of these plots is available in Appendix C.5.

Not only does the model match total production, but it also replicates changes in the fuel mix at varying demand levels, reflecting that plants differ in how balancing authorities dispatch them as a function of excess demand. Figure 5 shows the percentage of total production in each interconnection produced by plants of each fuel type in the model and the data. The X-axis of each panel varies the interconnection-level excess load—the total amount of electricity demand that must be satisfied by dispatchable plants. Across all interconnections, our model's predictions match the observed fuel mix very well. In all interconnections in the model and data, natural gas as a share of production increases in excess load. Meanwhile,

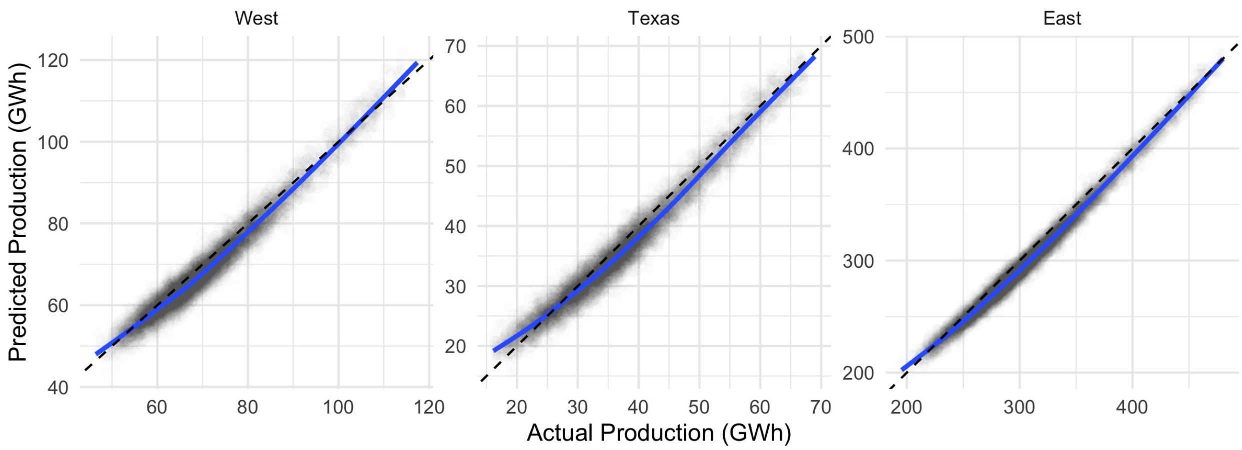


Figure 3: Model fit at the interconnection level. The Y-axis gives the production of dispatchable plants predicted by the model, measured in GWh while the X-axis gives the production in the data. Dots represent an hour of production for each interconnection, smoothed lines show the fit of a generalized additive model.

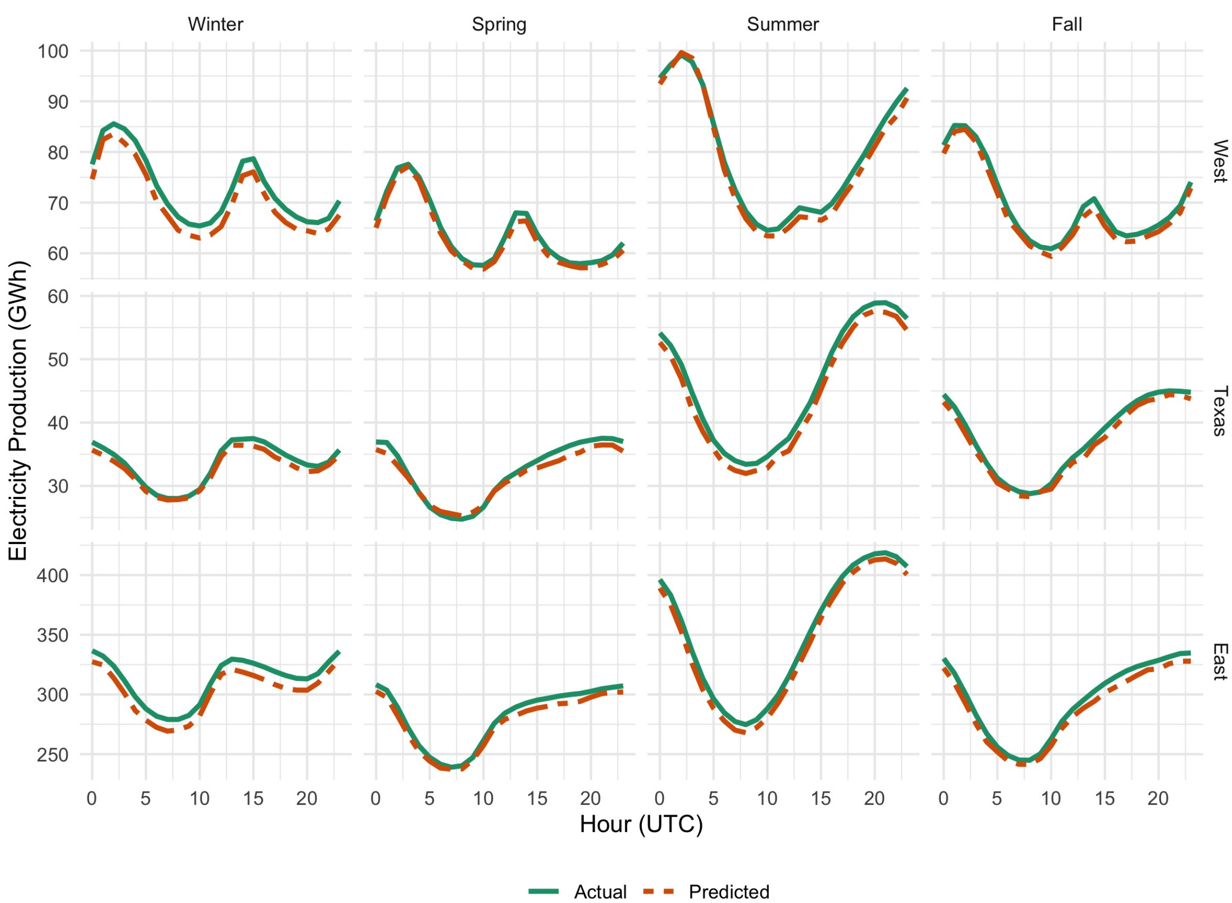


Figure 4: Model fit at the interconnection level by hour and season. Each panel shows predicted and actual production of dispatchable plants over the course of the average day, for each of the three intersections and across four seasons. The green solid line gives electricity production in the data while the red dotted line gives predicted production.

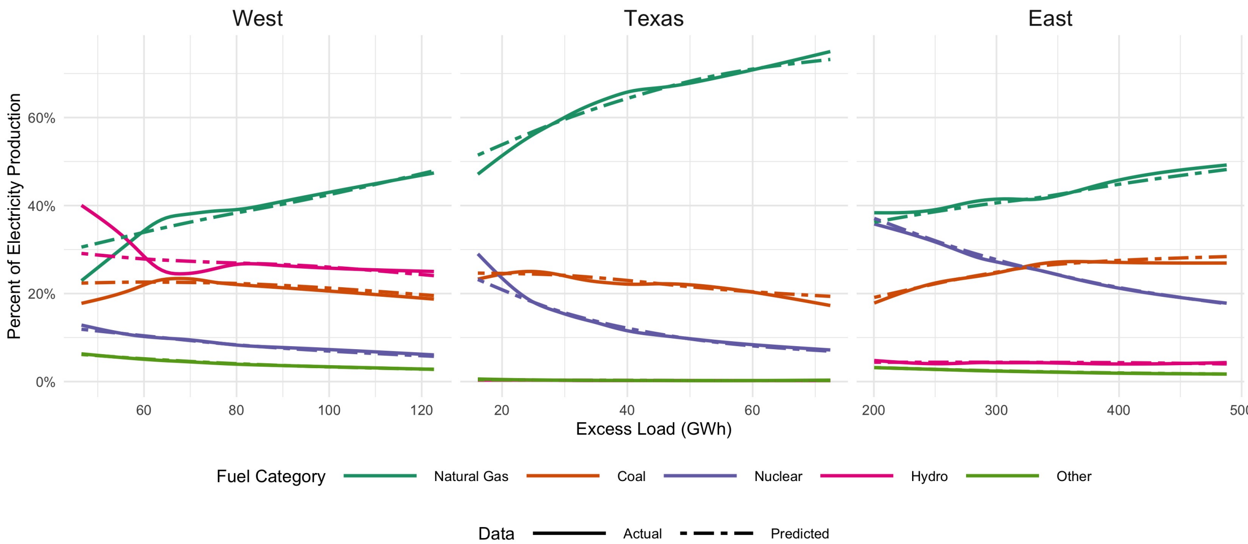


Figure 5: Fuel mix of production by interconnection. The X-axis gives excess load at the interconnection level and the Y-axis gives the percent of electricity production that is produced by each of the fuel types. The dashed lines show the fuel mix in the data while the solid lines show the simulated fuel mix.

production levels of clean, low-marginal-cost nuclear and hydroelectric plants generally decrease as a percentage of total production. An important difference between the Eastern and Western Interconnections is that coal increases its share of production in the East, whereas coal is declining in its share of production except at the lowest levels of excess load in the West.

These changes in the fuel mix imply that the marginal damages of electricity production may vary not only spatially but also as a function of electricity demand. To illustrate this, Figure 6 plots simulated marginal damages of energy production in each region within each interconnection as a function of excess load.[57] Overall, marginal damages are highest in regions within the Eastern interconnection, reflecting, in part, the interconnection's reliance on production from coal-fired power plants. However, there is significant heterogeneity in the marginal damages across regions within this interconnection. Marginal damages are highest from power plants the RFC region, which spans much of the Mid-Atlantic and lower Great Lakes.[58] Regions also vary in the extent to which their marginal damages of production change in excess load. In the Western and Texas interconnections, marginal damages are

[57]To calculate this, we simulate increasing excess load by a small amount in the region in question. We then divide the resulting change in total damages associated with power plants in the region by the change in total production by these power plants. Note that this is the marginal damage with respect to electricity production within a given region, not electricity demand from a given region. To the extent that a region imports electricity from other dirtier or cleaner regions, the marginal damage of electricity demanded may be higher or lower. For example, the NPCC imports electricity from the relatively dirty RFC, making marginal damage of electricity demand in NPCC higher than the electricity produced there.

[58]We include a map of NERC regions in Appendix A.7.

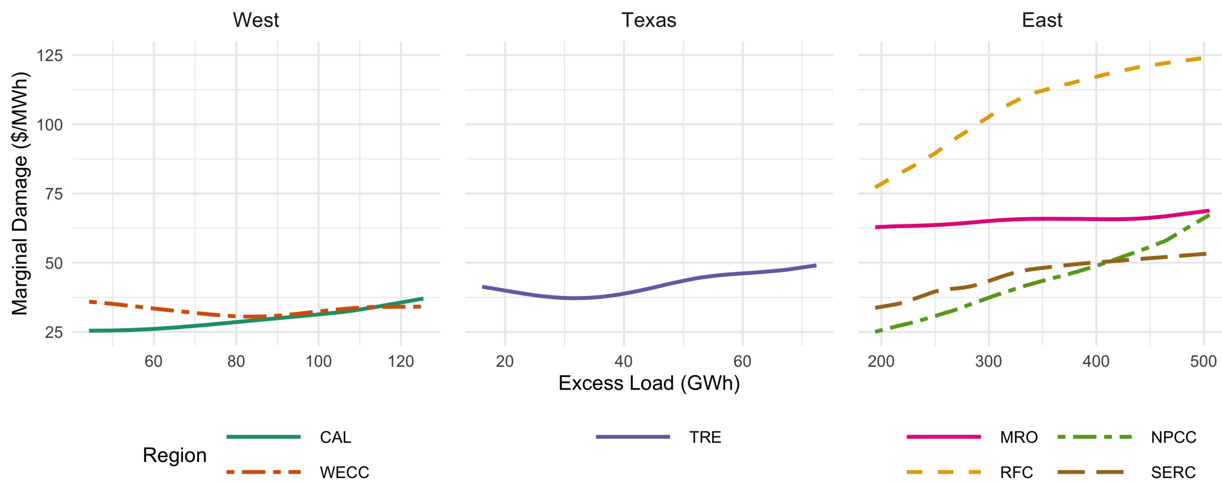


Figure 6: Estimated marginal damage of electricity production by region. The X-axis of each panel varies the total excess load in each of the three interconnections, and the Y-axis gives the simulated marginal damages per MWh of electricity produced in each region. See text for additional details.

relatively flat as a function of excess load while marginal damages in several regions in the Eastern interconnection are strongly increasing in excess load.[59] For example, in the NPCC, the region covering the Northeast, marginal damages increase by nearly 30% between the 25th to 75th percentile of excess load, going from $33/MWh to $43/MWh.

6 Counterfactuals and Optimal Subsidies

6.1 Welfare-Maximizing Subsidies

We now use the estimated structural model to quantify the welfare-maximizing solar subsidies characterized by equation (8).[60] The results are displayed in Figure 7 and in the first two columns of Table 2 and Table 3.

Before delving into how subsidies vary across states, we first analyze how the government should optimally allocate subsidies across the three subsidy types: cost-based subsidies, unit subsidies, and production-based (per-kWh) subsidies. To facilitate comparison, we calculate the present discounted value an "average installation" would receive. Specifically, we calculate the subsidy value every household in the model would receive if they purchased a 15-panel installation.[61] We then average this hypothetical subsidy value over all households. Table 2 shows the percent of the total subsidy value coming from each subsidy type in each simula-

[59] See Holland and Mansur (2008) for a discussion which types of power plants satisfy peak demand across regions and the implications for real-time electricity pricing.

[60] We outline the algorithm we use to numerically solve for welfare-maximizing subsidies in Appendix B.4.

[61] We define households as rooftops suitable for solar panel installations as defined by GPS data.

	(1)	(2)	(3)
		State-Specific Subsidies	
	Baseline	Welfare Max	Damage Min
Unit Subsidies	6.8	0.3	0.4
Cost Subsidies	81.6	0.9	0.9
kWh Subsidies	8.1	98.8	98.7
Total	100.0	100.0	100.0

Table 2: Percent of total subsidy value from each type of subsidy for a 15-panel installation averaged across all households in the model. Each column shows the subsidy values for a different simulation.

tion. Under the current system, over 80% of the value of subsidies comes from cost-based subsidies, which include the federal investment tax credit, state investment tax credits, sales tax exemptions, and property tax exemptions. On the other hand, the welfare-maximizing subsidies almost exclusively consist of production-based subsidies, with 98.8% of the value of subsidies coming from this type of subsidy. Intuitively, production-based subsidies incentivize installations for households where sunlight, and therefore environmental benefits, are high.[62] However, as we show in Section 6.4, the gains to reallocating across subsidy types within states are small relative to the gains from reallocating subsidies *across* states.

Next, Figure 7a and the first panel of Table 3 show how subsidy generosity varies across states under the optimal system of subsidies. We measure subsidy generosity as the present discounted value of subsidies a household would receive if they installed a 15-panel system.[63] Washington and Oregon, two states with relatively little sunlight and environmentally friendly power plants have the lowest optimal subsidies, at under 11 thousand dollars in present value. On the other end of the spectrum, five states in the RFC region have optimal subsidies valued at over 19 thousand dollars.[64] In West Virginia, one of these five states, current subsidy levels are some of the least generous in the country at under 6 thousand dollars. More generally, optimal subsidies are highest in the Mid-Atlantic and lower Great Lakes and are lowest in the Northwest.

Figure 7b and the second panel of Table 3 quantify the misallocation caused by the current system of subsidies on the spatial distribution of solar panel installations.[65] Current

[62] One important caveat is that we assume households' discount rate is given by the inverse of the real interest rate. De Groote and Verboven (2019) find that households use a much higher implicit interest rate than the market interest rate when evaluating the future benefits of solar panel installations. Therefore, upfront investment subsidies are more cost-effective than production subsidies at inducing installations. It would be straightforward to examine the robustness of our results to alternative values for the household discount rate.

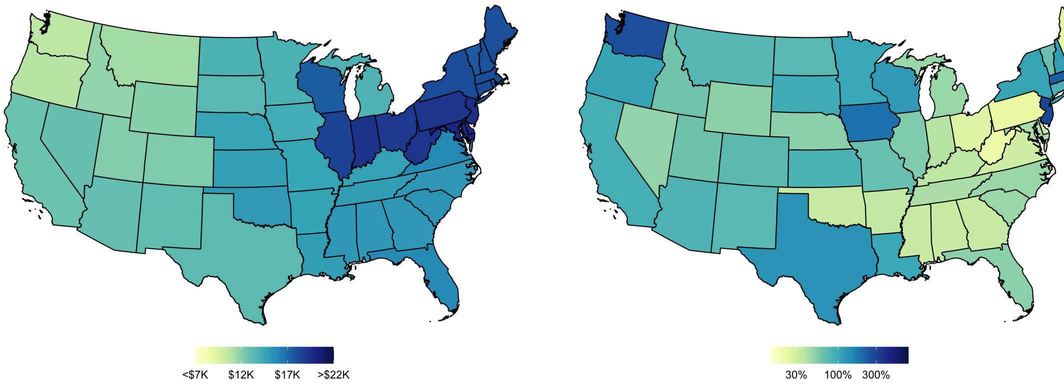
[63] This amount will vary within a given state, as per-kWh subsidies depend on the local level of solar irradiance. We take the population-weighted average across tracts within each state. We show state-level results in table form in Appendix C.7.

[64] These five states are Delaware, New Jersey, Indiana, West Virginia, and Pennsylvania.

[65] We find that almost all of the adjustment comes via the extensive margin, rather than the intensive

	(1)	(2)	(3)	(4)	(5)
		State-Specific Subsidies		Tract-Specific Subsidies	
	Baseline	Welfare Max	Damage Min	Welfare Max	Damage Min
I. Average Subsidy ($Thousands)					
Midwest	11.3	16.7	17.1	16.7	17.0
Northeast	18.2	18.6	21.8	18.6	21.6
South	10.4	15.0	12.8	15.1	12.7
West	12.0	12.4	5.7	12.4	5.6
II. Installations per 1000HHs					
Midwest	4.1	6.9	8.6	6.9	8.9
Northeast	29.8	20.2	28.5	20.2	27.9
South	6.8	11.1	9.6	11.1	9.6
West	13.2	14.0	6.7	13.9	6.6
National	11.4	12.4	11.4	12.4	11.3
III. Annual Damages Offset ($Millions)					
CO2e	69.8	75.9	72.6	75.9	72.5
NOx	18.0	19.3	18.3	19.3	18.2
PM2.5	16.7	16.9	18.1	16.9	18.1
SO2	37.0	38.3	48.5	38.8	49.4
Total	141.5	150.5	157.5	150.9	158.2

Table 3: Panel I shows the average present discounted value of subsidies received for a 15-panel installation for each census region. Panel II gives the simulated number of solar installations per 1000 households in the model for each Census region. Panel III gives the total damages offset by rooftop solar. See text for details on each simulation.



(a) Optimal subsidies. (b) Baseline installations as a percentage of optimal.

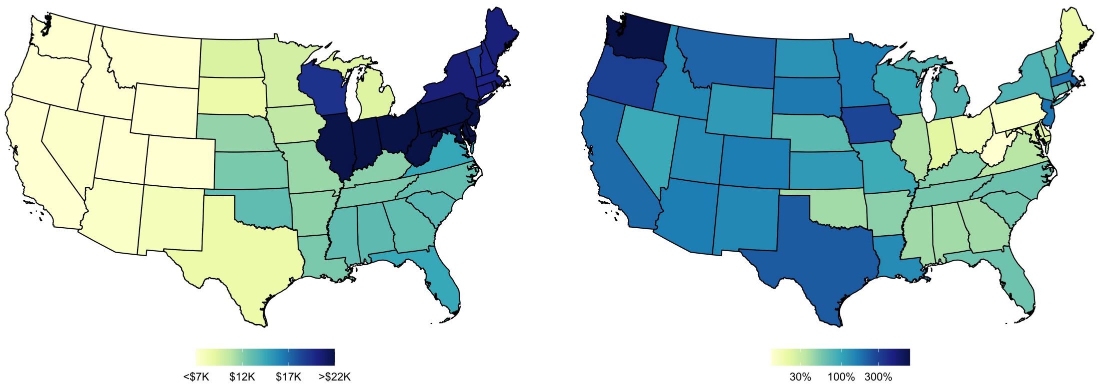
Figure 7: State-level optimal subsidies and misallocation for welfare-maximizing reforms. Panel (a) gives the optimal state subsidies. Subsidies are measured as the present discounted value associated with a 15-panel installation, averaged across all households in the state. Panel B shows state-level installations under the current system as a percentage of installations under the optimal system. These results are shown in table form in Appendix C.7.

installations in the Midwest and South are roughly 40% lower than under optimal subsidies, while installations in the Northeast are nearly 50% higher than the optimal level. Pennsylvania, for example, has only 23% of the optimal number of installations, while Massachusetts, New Jersey, and Washington have over twice as many installations as optimal. These results suggest that the current system of subsidies leads to a substantial misallocation of solar panels across states.

Panel III of Table 3 summarizes the environmental cost of this misallocation. Switching from the current subsidy scheme to the optimal scheme would increase annual damages offset by rooftop solar from \$141.5 million to \$150.5 million, equal to over a 6% increase in the aggregate environmental benefits of solar panels.[66] A decrease in CO2 equivalent emissions drives most of the environmental gains, with relatively minor effects on damages from other pollutants.

margin (number of panels per installation). In Appendix C.6, we show how the average installation size changes across counterfactuals.

[66]These environmental benefits are considerably smaller than the environmental benefits of reallocating panels found in Sexton et al. (2021). There are two main reasons for this difference. First, we consider a government with much more limited policy instruments. Sexton et al. (2021) consider a planner who can allocate panels across states freely, subject to a constraint that total installations are equal to the current level and that installations in a given zip code do not exceed the total number of detached single-family homes (or 30 percent of these homes). Here we consider a government which can only influence installations through subsidies. Second, we utilize emissions data from 2019 while Sexton et al. (2021) uses data from 2007-2016. Holland et al. (2020) find that power plant emissions decreased dramatically between 2010 and 2017. This decline was especially large in the Eastern interconnection, where emissions are historically higher than in the Western interconnection. This decrease in variation of damages across locations lowered the environmental benefits of reallocating panels across space.



(a) Optimal subsidies (b) Baseline installations as a percentage of optimal.

Figure 8: State-level optimal subsidies and misallocation for damage-minimizing reforms. Panel (a) gives the optimal state subsidies. Subsidies are measured as the present discounted value associated with a 15-panel installation, averaged across all households in the state. The color scale censors subsidy levels below $7K and above $22K. Panel B shows state-level installations under the current system as a percentage of installations under the optimal system. These results are shown in table form in Appendix C.7.

6.2 Damage-Minimizing Reforms

An alternative social objective is to choose the system of subsidies that minimizes environmental damages. In this section, we consider a government who chooses subsidies to minimize the net present value of environmental damages, $D\left(\text{ELoad}\right)$, subject to the government budget constraint. We formalize the government's problem and present the first-order conditions in Appendix B.5.

Figure 8 and the third columns of Tables 2 and Table 3 show the results. Like the welfare-maximizing policies, the damage-minimizing policies are most generous in the Mid-Atlantic, and are least generous in the Northwest. However, the variation across states in subsidy generosity is greater than under the welfare-maximizing subsidies: optimal damage-minimizing subsidies range from under 2 thousand dollars in Washington to over 24.5 thousands dollars in Maryland and Delaware. The reallocation of solar panels induced by the damage-minimizing subsidies would lead to approximately an 11% increase in aggregate environmental benefits.

6.3 Tract-level Subsidies

In the results above, we found that optimally set state-level subsidies lead to large environmental benefits relative to the current system of subsidies. Could subsidies set at a more granular geographic level lead to even larger gains? To answer this, we solve for the welfare-maximizing and damages-minimizing subsidies when subsidy levels are allowed to vary nonparametrically across census tracts.[67]

[67] A_i is constant within census tracts in our quantitative model, and therefore solar production within census tract is simply proportional the number of panels installed. Thus, there is no unique optimal system

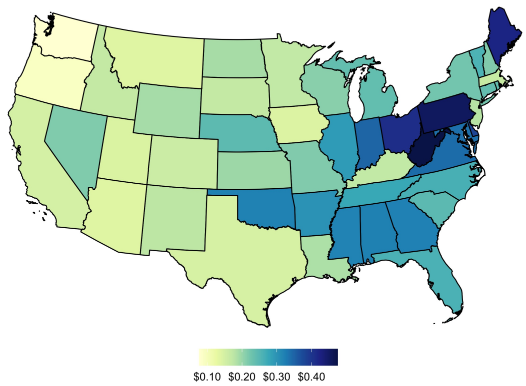


Figure 9: Damages offset per additional dollar of government funds associated with marginal increases in production-based subsidies, s_j^{kWh}, around the current system of subsidies. The results for other subsidy types are very similar and are presented in Appendix C.8.

Columns (4) and (5) of Table 3 display the results. In both the welfare-maximizing and damage-minimizing cases, the average subsidies across regions and the distribution of installations with optimal tract-level subsidies are similar to those under optimal state-level subsidies, and the damages offset with optimal tract-specific subsidies are only slightly larger than those with optimal state-specific subsidies. We conclude that the optimal system of state-level subsidies can capture nearly all of the gains of more geographically granular subsidies.

6.4 Marginal Subsidy Increases

Starting from the current system of subsidies, what marginal subsidy increases are the most cost-effective way to decrease damages? To answer this, we calculate the damages offset per additional dollar of government cost associated with marginal subsidy increases around the current system of subsidies.[68]

Figure 9 shows the marginal damages offset per dollar for production-based subsidies. There are large differences in damages offset across states. For example, a small subsidy increase in Washington only leads to roughly 8 cents less environmental damages per dollar

of subsidies when the planner can use both unit-based and production-based subsidies. Therefore, we set unit subsidies to 0 in this exercise.

[68]Using the notation introduced in Section 2.3, this is given by

$$\frac{\frac{\partial M_j}{\partial s_j^\theta}\overrightarrow{\Delta D}_j^{\theta,\text{ext}} + \left.\frac{\partial N_j}{\partial s_j^\theta}\right|_{M_j^{\text{st}}}\overrightarrow{\Delta D}_j^{\theta,\text{int}}}{\frac{\partial M_j}{\partial s_j^\theta}\overrightarrow{s}_j^{\theta,\text{ext}} + \left.\frac{\partial N_j}{\partial s_j^\theta}\right|_{M_j^{\text{st}}}\overrightarrow{\frac{\partial s}{\partial N}}_j^{\theta,\text{int}} + M_j\overline{\frac{\partial s_{ij}}{\partial s_j^\theta}}}.$$

of government funds. On the other hand, subsidy increases in West Virginia are highly cost-effective—for an additional dollar of government spending, environmental damages decrease by 48 cents. Subsidy increases in Ohio, Maine, and Pennsylvania are also associated with damages offset per dollar of over 40 cents. These results imply that small, cost-neutral shifts in subsidies across states could lead to substantial decreases in environmental damages. For example, a cost-neutral shift from subsidies in Washington to subsidies in West Virginia would lead to decreases in environmental damages of roughly $48 - 8 = 40$ cents for each dollar reallocated. Put another way: if the goal of Washington policymakers were to reduce total environmental damages, they would be significantly better off subsidizing solar installations in West Virginia, rather than in Washington.

We present the marginal damages offset for cost-based and unit-based subsidies in each state in Appendix C.8. Within each state, there are only small differences in the damages offset per dollar across the three subsidy types.

7 Extensions, Robustness and Further Issues

7.1 Alternative Specifications of Household Utility

In our baseline specification, we specified $\gamma_i(\cdot)$, the function which dictates a household's nonpecuniary benefits of solar installation, as a function of the number of panels installed and the local average education level and political leaning. In Appendix D.1, we assess the sensitivity of our results to this specification of the utility function by changing the specification of $\gamma_i(\cdot)$. In each specification, we re-estimate the model given the alternative specification of utility and then solve for the optimal policy given the new estimates of the household utility function. Across all specifications, we find similar optimal subsidies, similar changes in installation rates, and similar environmental benefits.

7.2 Line Losses

Our baseline model does not account for line loss: the electricity that is lost as electricity is transmitted over the grid from a power plant to a consumer. Rooftop solar reduces line loss by reducing the amount of electricity that needs to be transported across the grid.

In Appendix D.2, we re-calculate our main results in a model which accounts for line loss, where we base our model of line loss on the model and estimates from Borenstein and Bushnell (2022). In the extension, line losses are determined endogenously as a function of the amount of electricity in each region that must be transmitted between commercial plants and households. Therefore, residential solar offsets damages not only by directly producing power that would otherwise be produced by fossil-fuel plants, but also by reducing transmission across the grid and the resulting line losses.[69]

The takeaways are qualitatively the same as our main results. As expected, the environmental benefits solar panels increase. Switching to the welfare-maximizing and damage-minimizing subsidies generate increases in aggregate environmental benefits of 6.5% and 11.1%, respectively.

7.3 Improved Storage of Nondispatchable Technology

A significant issue facing the expansion of renewable electricity generation is that solar and wind are nondispatchable. Thus, these sources can only produce electricity when environmental conditions are suitable—when the sun is shining, or the wind is blowing. One of the leading solutions to this problem is an expanded capacity of electricity storage in the form of batteries. In Appendix D.3, we consider a stylized way to incorporate storage technology into our model. We allow nondispatchable electricity to be stored and used proportionately to the total load. Effectively this means we reallocate solar and wind production from their exogenous time profile of production to match the time profile of demand, which loosely matches the optimal behavior of storage owners arbitraging electricity across time to maximize profits. Adding storage technology does not qualitatively change the optimal subsidies, the distribution of installations under the optimal subsidies, nor the environmental benefits of switching to optimal subsidies. However, the storage technology itself generates considerable environmental benefits. See Butters, Dorsey, and Gowrisankaran (2021) or Holland, Mansur, and Yates (2022) for a detailed treatment of storage technology.

7.4 Cleaner Commercial Production

Electricity production in the United States has become considerably cleaner over the past few decades. Our baseline results quantify the value of optimizing solar panel subsidies given current electricity production technology. Here, we are interested in determining what would happen to our main results if the grid were considerably cleaner than it is presently.

Increased production of utility-scale renewables and fuel switching (from dirty to clean coal and from coal to natural gas) are the two primary drivers of the reduction in emissions from electricity generation. We perform four additional simulations to assess how further clean-up of commercial electricity production would affect our results. First, we find the optimal subsidies under expanded production from utility-scale solar and wind based on

[69]One caveat is that we do not adjust electricity production from solar panels to account for line losses when residential solar panels transmit electricity back into the grid.

three scenarios of projected renewable expansion by 2030 from the EIA (Nalley and LaRose, 2022).[70] Second, we recalculate results considering each coal plant to have "cleaned up" production. Our method of cleaning up coal plants is to adjust marginal damages from coal plants so that the mean and standard deviation of marginal damages from coal plants match that of natural gas plants.[71]

We present the results in Appendix D.4. We find that the damages offset by solar panels decrease with cleaner commercial production. However, we still find substantial benefits of switching from the current subsidies to optimal subsidies. Across all simulations, we find that switching to the optimal subsidies would lead to increases in damages offset of 5.2% to 10.8%.

7.5 Distributional Effects

The proposed switch in the system of subsidies could have distributional impacts through two channels—directly through a change in subsidies received by households and indirectly through the induced change in pollutant damages caused by electricity generation. Households who install solar panels, and therefore receive subsidies, tend to be wealthy (Borenstein and Davis, 2016). However, our proposed optimal subsidies will likely be progressive relative to the current subsidies since switching from current to optimal generally involves decreasing subsidies in higher-income states such as Massachusetts and increasing subsidies in lower-income states such as West Virginia. For this same reason, switching to optimal subsidies will likely lead to improvements in the distribution of damages caused by electricity generation. Similarly, Dauwalter and Harris (2023) find that shifting solar capacity to locations where the environmental benefits are greatest would lead to large environmental benefits for disadvantaged groups.

We have refrained from accounting for distributional effects when calculating optimal subsidies, as this paper is primarily concerned with the spatial misallocation of solar due to differences in the generosity of solar subsidies across states. Seriously tackling the distributional effects of solar subsidies requires a different set of policy instruments than those analyzed here, such as means-tested subsidies for solar installations.

[70] Specifically, we expand wind and solar based on their "reference case" projection, their low-cost projection, and their high-cost projection. We refer to their reference case projection as the mid-cost projection.

[71] This is admittedly ad-hoc and assumes that the dispatch curve does not change in response to the clean-up of coal plants.

8 Conclusion

In this paper, we used a structural model of solar panel demand and electricity production to calculate the optimal state-level system of subsidies for solar panels and to quantify the benefits of switching to such a system. Our main conclusions are that the current system of subsidies leads to severe misallocation of residential solar installations across space, and there would be substantial benefits to moving to the optimal subsidies. Switching to the optimal system of subsidies would lead to a 6-11% increase in the aggregate environmental benefits of rooftop solar.

Future work can extend our model to incorporate endogenous entry and exit of electricity generators, as in Holland, Mansur, and Yates (2022). In that case, residential solar subsidies could disincentive entry of new generators, which could be costly from an environmental perspective if the new generators employ cleaner technology than incumbents. It would also be interesting to allow for general equilibrium changes in electricity and solar panel prices. Future work could also utilize similar frameworks to analyze other consumer subsidies for energy-related products, such as subsidies for home insulation, small wind systems, and geothermal heat pumps. We leave these questions for future work.

References

Albouy, David. 2009. "The unequal geographic burden of federal taxation." *Journal of Political Economy*, 117(4): 635–667.

Anderson, Soren T, Ryan Kellogg, and James M Sallee. 2013. "What do consumers believe about future gasoline prices?" *Journal of Environmental Economics and Management*, 66(3): 383–403.

Arkolakis, Costas, and Conor Walsh. 2022. "Clean Growth." *Working Paper*.

Babalievsky, Fil, Kyle Herkenhoff, Lee E Ohanian, and Edward C Prescott. 2021. "The Sky is Not the Limit: The Macroeconomic Consequences of Commercial Zoning Laws."

Blair, N. J., A. P. Dobos, and P. Gilman. 2013. "Comparison of Photovoltaic Models in the System Advisor Model: Preprint." National Renewable Energy Lab. (NREL), Golden, CO (United States) NREL/CP-6A20-58057.

Bollinger, Bryan, Kenneth Gillingham, A Justin Kirkpatrick, and Steven Sexton. 2022. "Visibility and peer influence in durable good adoption." *Marketing Science*.

Borenstein, Severin, and James B. Bushnell. 2022. "Do Two Electricity Pricing Wrongs Make a Right? Cost Recovery, Externalities, and Efficiency." *American Economic Journal: Economic Policy*, 14(4): 80–110.

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.

Brandt, Loren, Trevor Tombe, and Xiaodong Zhu. 2013. "Factor market distortions across time, space and sectors in China." *Review of Economic Dynamics*, 16(1): 39–58.

Burr, Chrystie. 2014. "Subsidies, tariffs and investments in the solar power market." *University of Colorado-Boulder Working Paper*.

Butters, R. Andrew, Jackson Dorsey, and Gautam Gowrisankaran. 2021. "Soaking Up the Sun: Battery Investment, Renewable Energy, and Market Equilibrium."

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.

Chowdhury, Md Shahariar, Kazi Sajedur Rahman, Tanjia Chowdhury, Narissara Nuthammachot, Kuaanan Techato, Md Akhtaruzzaman, Sieh Kiong Tiong, Kamaruzzaman Sopian, and Nowshad Amin. 2020. "An overview of solar photovoltaic panels' end-of-life material recycling." *Energy Strategy Reviews*, 27: 100431.

Cicala, Steve, Stephen P Holland, Erin T Mansur, Nicholas Z Muller, and Andrew J Yates. 2021. "Expected health effects of reduced air pollution from COVID-19 social distancing." *Atmosphere*, 12(8): 951.

Colas, Mark, Sebastian Findeisen, and Dominik Sachs. 2021. "Optimal need-based financial aid." *Journal of Political Economy*, 129(2): 492–533.

Crago, Christine Lasco, and Ilya Chernyakhovskiy. 2017. "Are policy incentives for solar power effective? Evidence from residential installations in the Northeast." *Journal of Environmental Economics and Management*, 81: 132–151.

Dauwalter, Travis E, and Robert I Harris. 2023. "Distributional Benefits of Rooftop Solar Capacity." *Journal of the Association of Environmental and Resource Economists*, 10(2): 487–523.

De Groote, Olivier, and Frank Verboven. 2019. "Subsidies and time discounting in new technology adoption: Evidence from solar photovoltaic systems." *American Economic Review*, 109(6): 2137–72.

Fajgelbaum, Pablo D, Eduardo Morales, Juan Carlos Suárez Serrato, and Owen Zidar. 2019. "State taxes and spatial misallocation." *The Review of Economic Studies*, 86(1): 333–376.

Feger, Fabian, Nicola Pavanini, and Doina Radulescu. 2022. "Welfare and redistribution in residential electricity markets with solar power." *The Review of Economic Studies*, 89(6): 3267–3302.

Gillingham, Kenneth, and Tsvetan Tsvetanov. 2019. "Hurdles and steps: Estimating demand for solar photovoltaics." *Quantitative Economics*, 10(1): 275–310.

Graff Zivin, Joshua S, Matthew J Kotchen, and Erin T Mansur. 2014. "Spatial and temporal heterogeneity of marginal emissions: Implications for electric cars and other electricity-shifting policies." *Journal of Economic Behavior & Organization*, 107: 248–268.

Heath, Garvin A, and Margaret K Mann. 2012. "Background and reflections on the life cycle assessment harmonization project." *Journal of Industrial Ecology*, 16(NREL/JA-6A20-54177).

Herkenhoff, Kyle F, Lee E Ohanian, and Edward C Prescott. 2018. "Tarnishing the golden and empire states: Land-use restrictions and the US economic slowdown." *Journal of Monetary Economics*, 93: 89–109.

Holland, Stephen P, and Erin T Mansur. 2008. "Is real-time pricing green? The environmental impacts of electricity demand variance." *The Review of Economics and Statistics*, 90(3): 550–561.

Holland, Stephen P, Erin T Mansur, and Andrew J Yates. 2021. "The electric vehicle transition and the economics of banning gasoline vehicles." *American Economic Journal: Economic Policy*, 13(3): 316–44.

Holland, Stephen P, Erin T Mansur, and Andrew J Yates. 2022. "Decarbonization and Electrification in the Long Run." National Bureau of Economic Research.

Holland, Stephen P, Erin T Mansur, Nicholas Z Muller, and Andrew J Yates. 2016. "Are there environmental benefits from driving electric vehicles? The importance of local factors." *American Economic Review*, 106(12): 3700–3729.

Holland, Stephen P, Erin T Mansur, Nicholas Z Muller, and Andrew J Yates. 2020. "Decompositions and policy consequences of an extraordinary decline in air pollution from electricity generation." *American Economic Journal: Economic Policy*, 12(4): 244–74.

Holland, Stephen P, Erin T Mansur, Nicholas Z Muller, and Andrew J Yates. 2021. "The environmental benefits of transportation electrification: Urban buses." *Energy policy*, 148: 111921.

Holland, Stephen P, Matthew J Kotchen, Erin T Mansur, and Andrew J Yates. 2022. "Why marginal CO2 emissions are not decreasing for US electricity: Estimates and implications for climate policy." *Proceedings of the National Academy of Sciences*, 119(8): e2116632119.

Hsieh, Chang-Tai, and Enrico Moretti. 2019. "Housing Constraints and Spatial Misallocation." *American Economic Journal: Macroeconomics*, 11(2): 1–39.

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.

Jacobsen, Mark R, James M Sallee, Joseph S Shapiro, and Arthur A van Benthem. 2022. "Regulating Untaxable Externalities: Are Vehicle Air Pollution Standards Effective and Efficient?"

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

Kirkpatrick, A Justin, and Lori S Bennear. 2014. "Promoting clean energy investment: An empirical analysis of property assessed clean energy." *Journal of Environmental Economics and Management*, 68(2): 357–375.

Kwan, Calvin Lee. 2012. "Influence of local environmental, social, economic and political variables on the spatial distribution of residential solar PV arrays across the United States." *Energy Policy*, 47: 332–344.

Lamp, Stefan, and Mario Samano. 2023. "(Mis)allocation of Renewable Energy Sources." *Journal of the Association of Environmental and Resource Economists*, 10(1): 195–229.

Langer, Ashley, and Derek Lemoine. 2022. "Designing dynamic subsidies to spur adoption of new technologies." *Journal of the Association of Environmental and Resource Economists*, 9(6): 1197–1234.

Liski, Matti, and Iivo Vehviläinen. 2020. "Gone with the wind? An empirical analysis of the equilibrium impact of renewable energy." *Journal of the Association of Environmental and Resource Economists*, 7(5): 873–900.

Miller, Gregory, Gailin Pease, Milo Knowles, and Wenbo Shi. 2022. "Open Grid Emissions Initiative."

Miller, Keaton S, Amil Petrin, Robert Town, and Michael Chernew. 2019. "Optimal managed competition subsidies." National Bureau of Economic Research.

Muller, Nicholas Z, Robert Mendelsohn, and William Nordhaus. 2011. "Environmental accounting for pollution in the United States economy." *American Economic Review*, 101(5): 1649–75.

Nalley, Stephen, and Angelina LaRose. 2022. "Annual energy outlook 2022 (AEO2022)." *Energy Information Agency*, 23.

Pigou, Arthur C. 1920. "The economics of welfare Macmillan and Co." *London, United Kingdom.*

Pretnar, Nick, and Alexander Abajian. 2021. "An Aggregate Perspective on the Geo-spatial Distribution of Residential Solar Panels." *Available at SSRN 3771496.*

Rennert, Kevin, Frank Errickson, Brian C Prest, Lisa Rennels, Richard G Newell, William Pizer, Cora Kingdon, Jordan Wingenroth, Roger Cooke, Bryan Parthum, et al. 2022. "Comprehensive evidence implies a higher social cost of CO2." *Nature*, 610(7933): 687–692.

Sarzynski, Andrea, Jeremy Larrieu, and Gireesh Shrimali. 2012. "The impact of state financial incentives on market deployment of solar technology." *Energy Policy*, 46: 550–557.

Sengupta, Manajit, Yu Xie, Anthony Lopez, Aron Habte, Galen Maclaurin, and James Shelby. 2018. "The National Solar Radiation Data Base (NSRDB)." *Renewable and Sustainable Energy Reviews*, 89: 51–60.

Sexton, Steven, A Justin Kirkpatrick, Robert I Harris, and Nicholas Z Muller. 2021. "Heterogeneous solar capacity benefits, appropriability, and the costs of suboptimal siting." *Journal of the Association of Environmental and Resource Economists*, 8(6): 1209–1244.

Shapiro, Joseph S. 2021. "The environmental bias of trade policy." *The Quarterly Journal of Economics*, 136(2): 831–886.

Shapiro, Joseph S, and Reed Walker. 2020. "Is Air Pollution Regulation Too Stringent?" National Bureau of Economic Research.

Siler-Evans, Kyle, Inês Lima Azevedo, and M Granger Morgan. 2012. "Marginal emissions factors for the US electricity system." *Environmental science & technology*, 46(9): 4742–4748.

Sodhi, Manbir, Lennart Banaszek, Chris Magee, and Mercedes Rivero-Hudec. 2022. "Economic Lifetimes of Solar Panels." *Procedia CIRP*, 105: 782–787.

Xu, Yan, Jinhui Li, Quanyin Tan, Anesia Lauren Peters, and Congren Yang. 2018. "Global status of recycling waste solar panels: A review." *Waste Management*, 75: 450–458.

Yu, Jiafan, Zhecheng Wang, Arun Majumdar, and Ram Rajagopal. 2018. "DeepSolar: A machine learning framework to efficiently construct a solar deployment database in the United States." *Joule*, 2(12): 2605–2617.

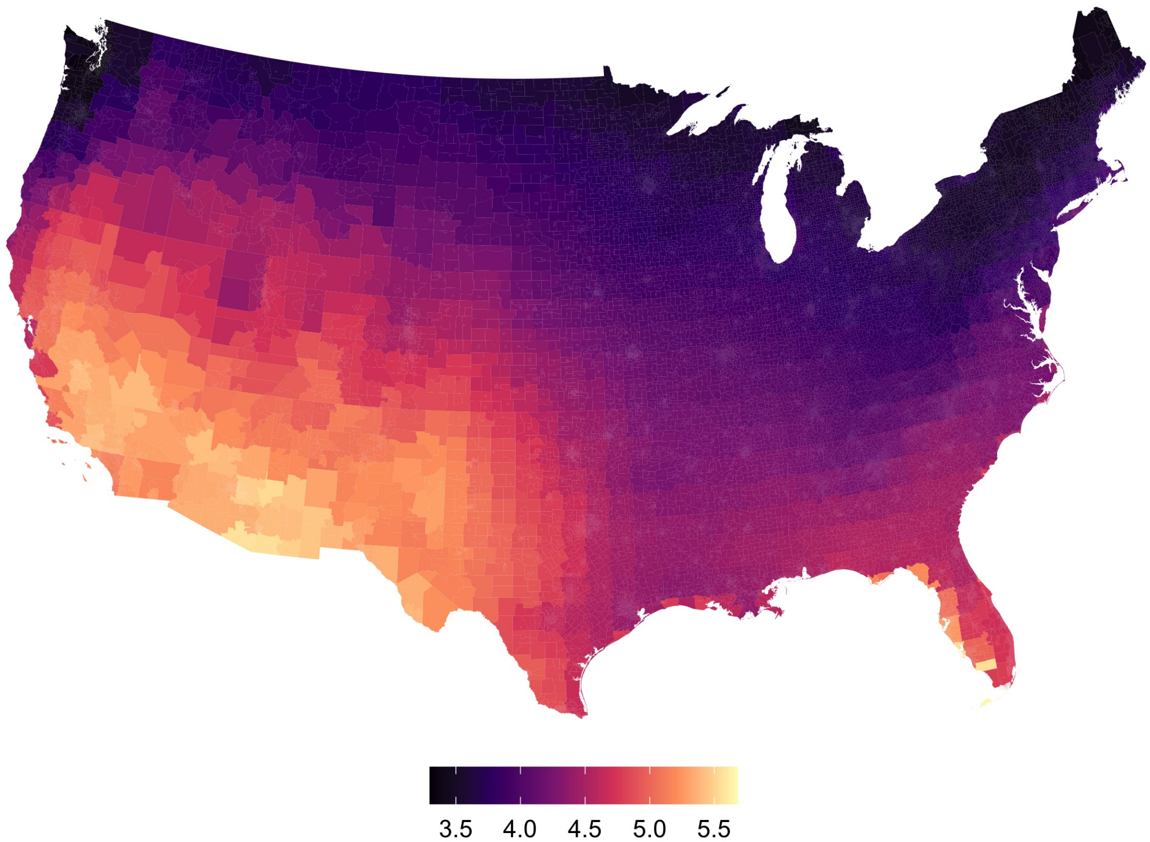


Figure A1: Daily solar radiation $\left(kWh/m^2/day\right)$ by census tract from Deepsolar.

A Data Appendix

A.1 Deep Solar

We obtain data on solar panel installation from the Deepsolar database, which is created by applying a novel semi-supervised deep-learning framework to satellite imagery from Google Static Maps from the year 2016 (Yu et al., 2018). The Deepsolar model predicts the number and size of solar panel installations across the contiguous United States. We use these tract-level data on the number and size of residential solar systems to give us our empirical analogs of BI_t and installation size N_i.

Deepsolar also estimates the daily solar radiation in each census tract, measured in kWh per square meter per day, which we show in Figure A1. For any missing tracts in the Deepsolar data, we impute daily solar radiation by taking the simple mean of any bordering tract with non-missing values.

A.2 Google Project Sunroof

For data on solar irradiance, A_i, and number of available spaces for panels $\bar{N}_i$, we utilize tract-level data from Google Project Sunroof (GPS). GPS begins with satellite imagery from Google Maps. It then applies a deep-learning algorithm to create 3D models of rooftops. These 3D models allow GPS to estimate the amount of sunlight a given rooftop receives over the course of the year, taking into account changes in the position of the sun over the course of the day and year. These 3D models are used to calculate the amount of available space for solar panels.

We assume that all households within a given tract have access to the same solar irradiance, which we measure as total solar energy generation potential for the average panel in a given tract. For number of potential panels $\bar{N}_i$, the GPS data provide then number of buildings in each tract with differing amounts of space available for solar panel installations. This effectively gives us the full distribution of $\bar{N}_i$ for households within a given tract.

One potential issue with the GPS data is that it might also capture potential space for solar panels that is not suitable for residential solar (for example not being part of someone's house). To deal with this, we limit potential solar sites in Google Project Sunroof to those with space available to 42 MW of solar panels, corresponding to the 99.9th percentile of the largest solar panel in the Tracking the Sun data described in Appendix A.3. The results are not sensitive to this censoring.

A.3 Tracking the Sun

Tracking the Sun is an aggregation of solar system installation data created by the Lawrence Berkeley National Lab. The Lawrence Berkeley National Lab collects these data from existing public databases and directly from state agencies, utilities, and other organizations. The result is 2.5 million solar installations from the last two decades, with installation price, system size, and subsidies geographically identified at the zip code level, along with other information about the installed solar system. The installations cover nearly 80 percent of all installed solar systems in the U.S. but include only 25 states. Some of these states do not include price data for any installations. Across all states, about a quarter of observations for residential solar system installations are missing price data.

We use the Tracking the Sun data to estimate prices for solar systems, using total cost and number of panels installed to estimate a fixed cost of installation and variable, per-panel cost. Since many states have no data, we assume pricing functions are common within each census region. We filter the Tracking the Sun data to include residential installations between 2000 and 2020 that are not missing price or the total number of panels, which leaves us with nearly 1.3 million observations. Additionally, we censor installation costs at the 0.5th and 99.5th percentiles and convert them into 2019 dollars.

A.4 System Advisor Model

While we obtain annual electricity generation for solar panels from Google Project Sunroof, those data do not include any information on how that production varies by hour within a year. Thus, we use the System Advisor Model (SAM) from the National Renewable Energy Laboratory to estimate hourly electricity profiles for each state (Blair, Dobos, and Gilman, 2013). SAM is an open-source program that estimates the performance of solar systems and other renewable power systems. We follow the methodology in (Sexton et al., 2021), calculating electricity generation for a system with typical parameters where tilt matches latitude and panels point south. The only difference is that we estimate generation for systems at the centroid of each state. We use weather data from Sengupta et al. (2018) for a state's typical meteorological year.

The model's output is the hourly production of a solar system over the course of a year in each state. We create hourly profiles by dividing the hourly generation by each state's total annual generation. Figure A2 for hourly production for examples of the results for six states.

A.5 State Electricity Prices

Figure A3 presents the state-level electricity prices we use in our empirical analysis. California and states in the Northeast have the highest electricity prices at over 15 cents per kWh. Most of the country has prices between 8 and 10 cents per kWh.

A.6 Subsidies

We calculate s_j^{kWh} as the sum of per-kWh rebates and the average price of Solar Renewable Energy Certificates. In some states (e.g. Massachusetts and New Jersey), households can only sell Solar Renewable Energy Certificates for a certain number of years after installation. For these states, we only calculate the value of Solar Renewable Energy Certificates for years in which households are permitted to sell the credits. We calculate s_j^{Cost} as the sum of federal investment tax credits, state investment tax credits, sales tax excemptions, and the net present value of property tax excemptions. We translate per-KW rebates to s_j^{Unit} by assuming a constant 0.25 KW per panel. Maryland has a fixed rebate of \$1000 per system. We translate this into a per-unit subsidy by dividing this amount by the averge number of panels in an installation (15). Many states place a cap on the maximum amount of a type of

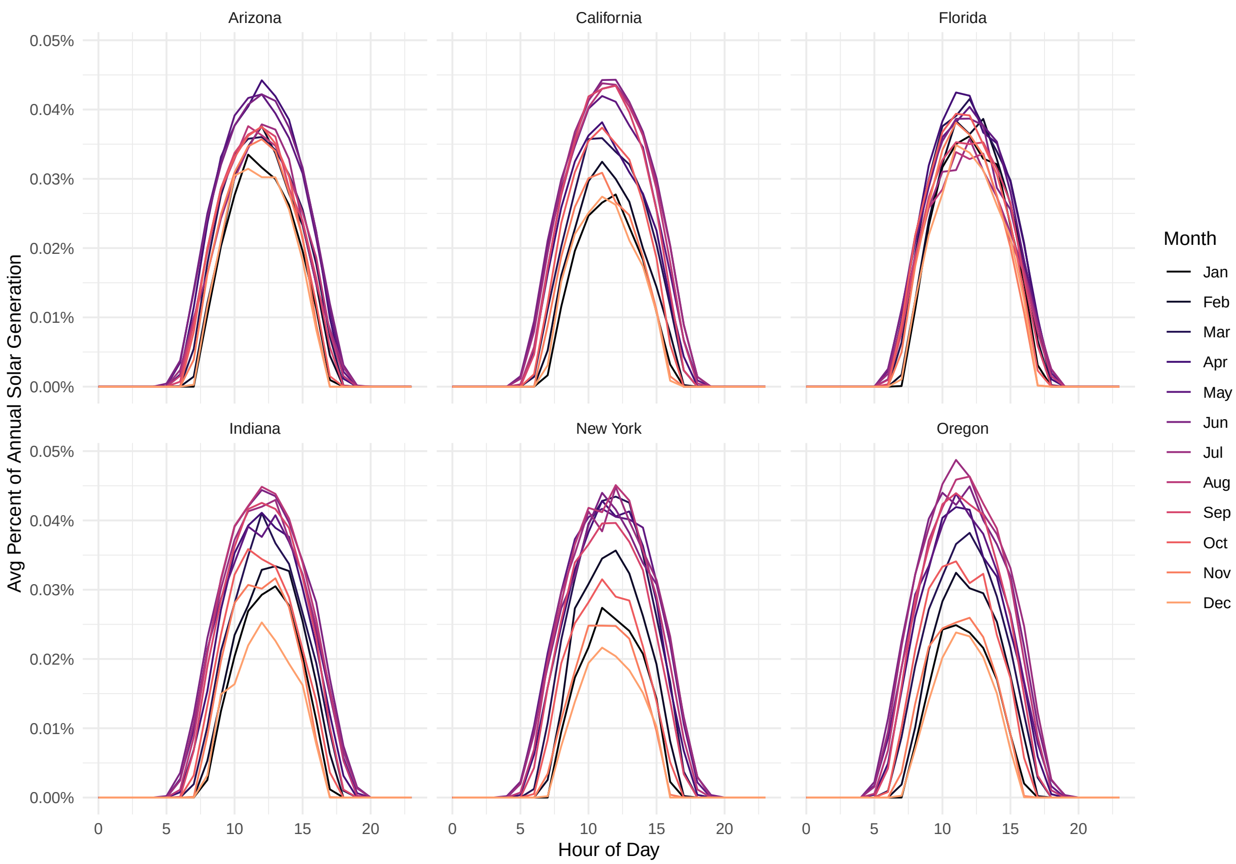


Figure A2: Hourly electricity generation for a standard solar panel for six example states.

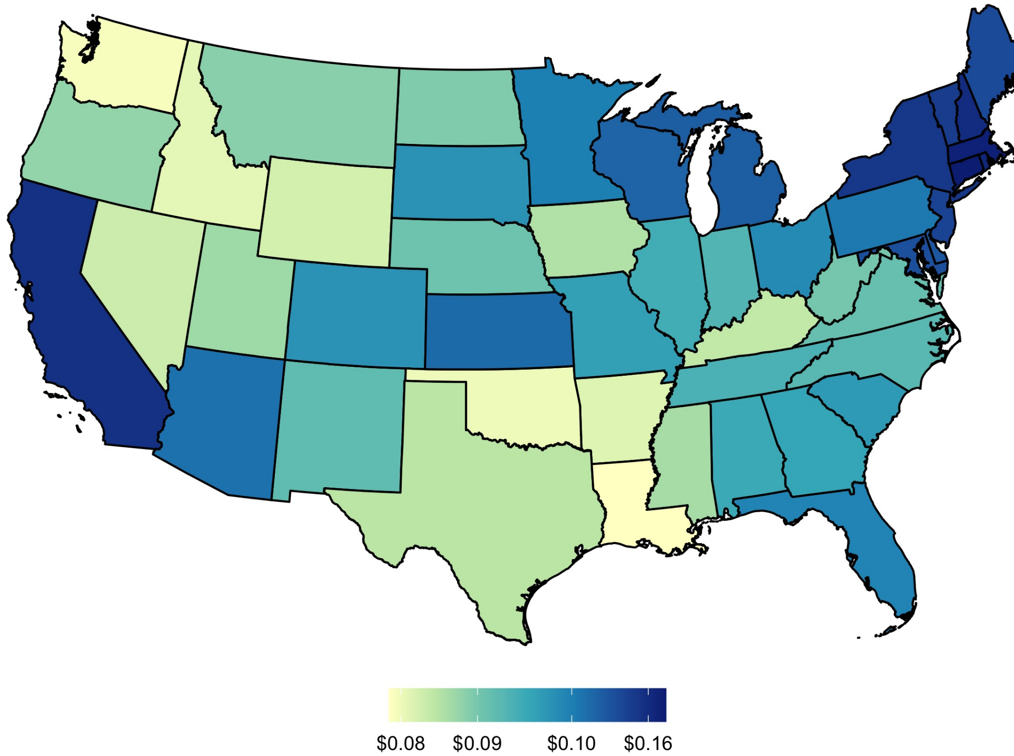


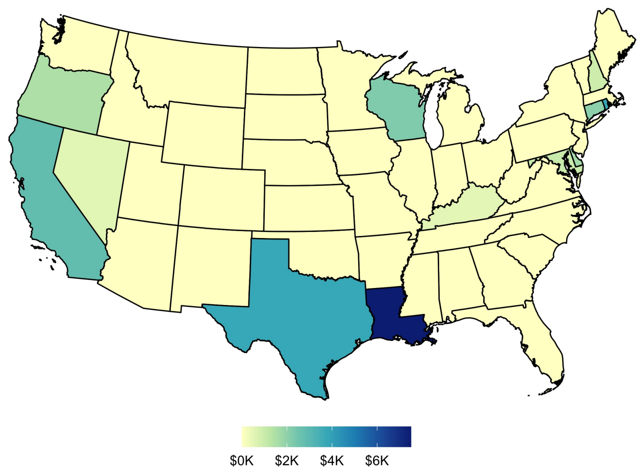
Figure A3: State electricity prices ($\$/kWh$)

subsidy a household can receive. We enforce these state-level maxima in estimation. Figure A4 shows the state level expected subsidies for a 15-panel system for each type of subsidy: unit, cost, and kWh. Most of the current subsidies take the form of cost-based subsidies, while few states offer kWh and unit-based subsides.

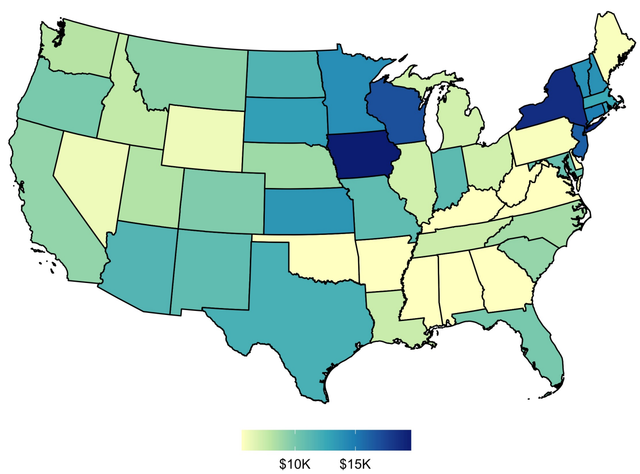
A.7 Power Plants

The following describes how we use the Open Grid Emissions (OGE) data. These data have several advantages over the commonly used raw electricity data from the EIA and EPA, which we describe below.

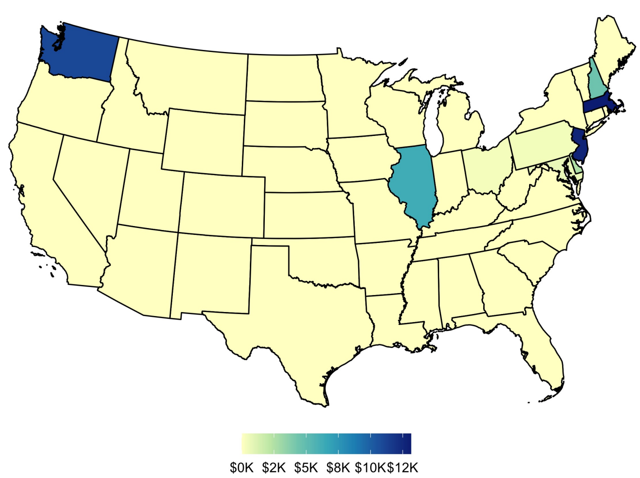
Plant level The hourly, plant-level data from OGE give net electricity generation and emissions of SO2, NOx, CO2, and CO2 equivalent. OGE derives these data primarily from the EPA CAMD, which reports hourly gross electricity generation and emissions at the unit level, where units typically correspond to generators connected to a single emissions stack. OGE adjusts gross generation to account for electricity losses before entering the grid and aggregates these units to the facility level, which we refer to as power plants. Additionally, OGE removes the portion of emissions from fuel burned for heat for combined heat and



(a) Unit-based subsidies



(b) Cost-based subsidies



(c) kWh-based subsidies

Figure A4: Expected subsidy for a 15-panel system by subsidy type power plants. OGE's static plant attributes table has a latitude and longitude for each plant, allowing us to match each power plant to a county.

We also collect each plant's nameplate capacity and stack height. Nameplate capacities are from EIA Form 860 and represent the maximum amount of electricity that a generating unit is rated to produce. We sum the nameplate capacities of generators in a plant to calculate plant-level nameplate capacities. Thirty-seven plants are missing from the EIA 860 data, for which we use nameplate capacities from the EPA's eGRID files from 2019 and 2020. We obtain stack heights from the EPA CAMD and set a plant's stack height as the median stack height of units the within that plant. Thirty plants are missing from these data, for which we set the stack height equal to the median stack height of all plants of the same primary fuel category. We use a power plant's stack height and location to assign the marginal damage coefficient in the AP3 model. We opt for CO2e over CO2 when calculating damages as it includes emissions of the more potent greenhouse gasses methane and nitrous oxide in addition to CO2.[72]

The EPA CAMD hourly unit-level data only include fossil-fuel plants with greater than 25 MW of generating capacity, leaving a non-negligible portion of generation and emissions unreported. One of the main goals of OGE is to ensure complete coverage of the electricity generation sector. In essence, they combine the reported hourly plant-level data from the EPA CAMD with hourly balancing authority-fuel category level data from the EIA to calculate a 'residual' profile, the unreported production from small or non-fossil-fuel power plants.

There are 9,514 plants with hourly production and emissions in the OGE data. About a third of the plants do not have observations for every hour in 2019. We fill in any generation and emissions values between the first and the last hour a plant appears in the data with zeros. After removing 4,388 non-dispatchable plants (wind and solar), 471 plants with zero or negative reported net electricity generation, and 30 plants with no variation in net electricity generation, we have 4,625 power plants—giving us 40.1 million plant-hour observations.

Emissions rates OGE's hourly data does not include PM2.5 emissions, as the EPA CAMD and the EIA do not report PM2.5 emissions from power plants. As a part of the eGRID project, the EPA has collected annual PM2.5 emissions from the National Emissions Inventory (NEI) and matched those emissions to electricity-generating units to calculate an average PM2.5 emissions rate. We match these estimated annual rates to our power plants, taking the production-weighted average over units within a power plant. We use the fuel category median value for the power plants missing PM2.5 emissions rates. This imputation

[72]As detailed in the OGE documentation, they calculate CO2e using the global warming potential of each GHG according to the IPCC's 5th Assessment Report. They calculate methane and nitrous oxide emissions using a constant, fuel-specific emissions factor.

Table A1: Summary statistics on dispatchable power plants by region

		Net Generation		Emissions (lb/MWh)		
Region	Number of plants	Total (TWH)	% Fossil Fuel	NOx	SO2	CO2e
CAL	639	1,611	58.8	0.67	0.07	670
MRO	1,182	8,006	79.6	0.99	1.10	1,356
NPCC	723	2,250	45.8	0.42	0.06	550
RFC	657	7,714	61.5	0.57	0.68	900
SERC	582	9,167	67.0	0.51	0.36	893
TRE	179	3,273	87.0	0.66	0.71	1,095
WECC	663	4,639	60.7	0.79	0.44	1,056
Total	**4,625**	**36,660**	**67.9**	**0.68**	**0.60**	**1,003**

only applies to 10% of total electricity production. Additionally, we censor PM 2.5 emissions rates at the 95th percentile for each fuel category.

Regions We follow Holland et al. (2016) in our definitions of regions for the electricity generation model. OGE assigns plants to the balancing authority in charge of dispatching the plant. We then assign balancing authorities to regions. There are six NERC regions in the contiguous US. Four of these (MRO, RFC, NPCC, and SERC) fall within the Eastern Interconnection, while the other two (WECC and TRE) are in the Western and Texas Interconnections, respectively.

Most BAs fall entirely within one NERC region, but some BAs have generating units in multiple NERC regions. For all BAs except MISO and PJM, we assign the BA to the NERC region with the most overlapping generating units between the BA and NERC region using the static plant attributes data from OGE. We assign the MISO BA to the MRO NERC region and the PJM BA to the RFC NERC region. Finally, we give California its own NERC region, consisting of five BAs: BANC, CISO, IID, LDWP, and TIDC. Figure A5 shows a map of these regions. We used the eGRID power profiler to assign approximate service areas for each region.

Table A1 shows summary statistics describing generation and average emissions in each region, highlighting the heterogeneity in average emissions between regions. This is largely driven by differences in the fuel mix between regions. Table A2 shows a summary of generation and emissions by fuel category.

Excess Load We calculate the excess load (total demand minus production from nondispatchable generating units) within each region using OGE's power sector-level data. These data give hourly net generation by fuel category for each balancing authority. We perform

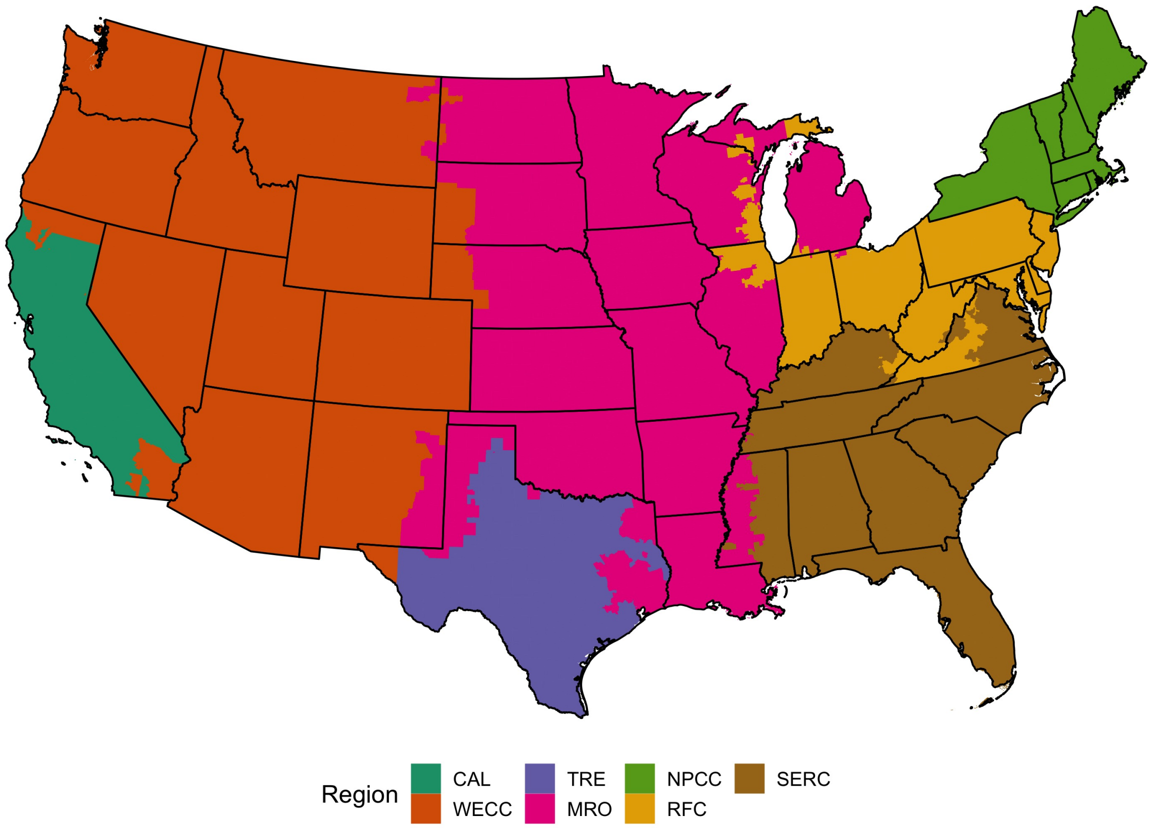


Figure A5: Map of region service areas

Table A2: Summary statistics on dispatchable power plants by fuel category

			Emissions (lb/MWh)		
Fuel	Number of plants	Net generation (TWH)	NOx	SO2	CO2e
Biomass	583	547	5.51	1.12	1,608
Coal	262	8,973	1.53	2.14	2,216
Natural Gas	1,696	15,905	0.44	0.06	934
Nuclear	60	8,074	0.00	0.00	5
Petroleum	535	26	19.04	40.59	22,506
Geothermal	60	155	0.00	0.34	136
Hydro	1,325	2,802	0.00	0.00	0
Other	38	45	0.04	0.00	98
Waste	66	133	5.89	1.34	3,747
Total	**4,625**	**36,660**	**0.68**	**0.60**	**1,003**

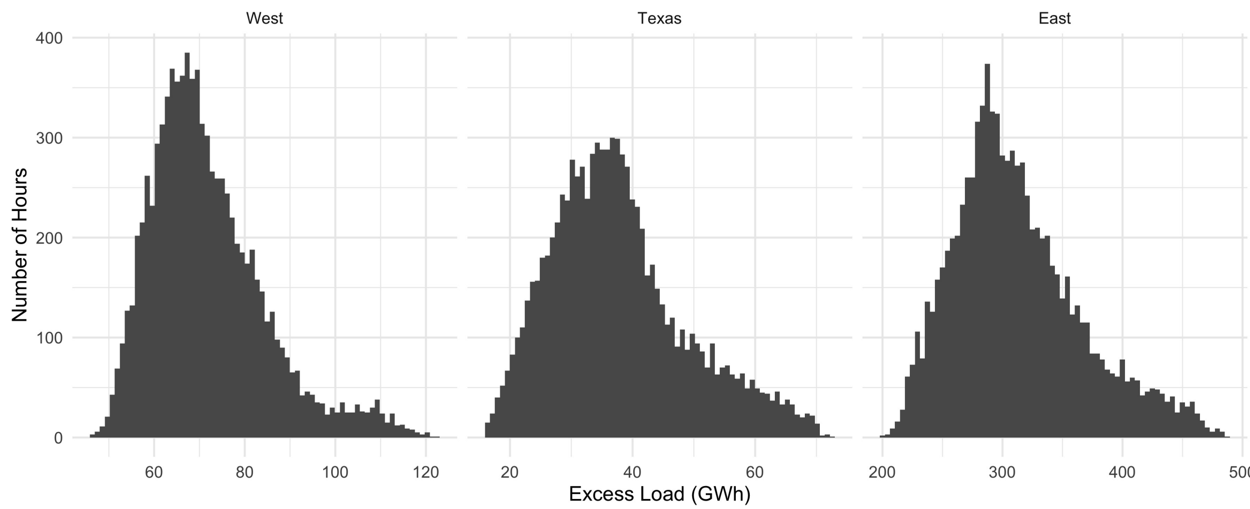


Figure A6: Histogram of total excess load within each interconnection. Observations are hours.

minor data cleaning to ensure that misreporting in the underlying data does not impact our estimates. We consider values above 1.5 times the 99th percentile for that balancing authority and fuel category to be outliers. We replace outliers with the value from the previous hour as long as the previous hour's value is not also an outlier. If the previous hour is an outlier, then we use the value from the same hour in the previous day as long as that is not an outlier. If the previous hour and the previous day are outlier values, we censor to 1.5 times the 99th percentile.[73] We then calculate the total load within a region as the sum of net generation across all balancing authorities and fuel categories within a region. Excess load is the total load in a region minus net generation from solar and wind, the two nondispatchable energy sources. Figure A6 shows the distribution of total excess load within each interconnection.

B Theory and Quantitative Appendix

B.1 States Without Net Metering

In the general model above, we assumed that households could sell back electricity produced by their solar panels at price p_j. This is the case if the state offers net-metering, which is offered in all but 9 in our sample.[74] In the states were net metering is not offered, states can sell back electricity to the grid at price $p_j^{sale} \leq p_j$. Let $A_i^{home}\left(N_i, \{e_{it}\}_{t=0}^{T}, \{A_{it}\}_{t=0}^{T}\right)$ give the discounted sum of energy that is used at home, written as a function of panels installed,

[73]This process mimics that used by the EIA when aggregating net generation by balancing authority to the region level, see the "Net Generation" section on this page.

[74]Idaho, Tennessee, Texas and Alabama do not have statewide mandatory net-metering policies. Idaho does not have state net-metering policy but each of the state's three investor-owned utilities have a net-metering policy. Five other states in our sample have distributed generation rules other than net-metering.

electricity consumption, and the stream of solar irradiance. Let $A_i^{grid}\left(N_i, \{e_{it}\}_{t=0}^T, \{A_{it}\}_{t=0}^T\right)$ be the discounted sum of energy that is sold back to grid, such that $A_i^{home}(\cdot) + A_i^{grid}(\cdot) = A_i$. We can write the budget constraint for households in states without net metering as

$$c + \underbrace{p_j\left(e - m_i N_i A_i^{home}(\cdot)\right)}_{\text{Cost of electricity}} + m_i \underbrace{\left[\left(1 - s_j^{\text{Cost}}\right) p_j^{\text{Ins}}(N_i)\right]}_{\text{Net cost of installation}} =$$

$$y_i + m_i\left(\underbrace{N_i A_i s_j^{\text{kWh}}}_{\text{kWh Subsidy}} + \underbrace{N_i s_j^{\text{Unit}}}_{\text{Unit Subsidy}} + \underbrace{p_j^{sale} N_i A_i^{grid}(\cdot)}_{\text{Electricity sold to grid}}\right). \quad (13)$$

In estimating and simulating the model, we assume that the household's optimal electricity consumption, $\{e_{it}^\star\}_{t=0}^T$, is independent of the household's installation decision. Again letting $N_i^\star$ represent the optimal choice of panels, we can then summarize the decision for households in states without met metering as

$$V_i = \max_{N_i, m_i \in \{0,1\}} + m_i\left[\hat{\mu}_{ij}(N_i) + \gamma_i(N_i^\star)\right],$$

where

$$\hat{\mu}_{ij}(N_i) = N_i A_i\left(\hat{p}_j + s_j^{\text{kWh}}\right) - \left(1 - s_j^{\text{Cost}}\right) p_j^{\text{Ins}}(N_i) + N_i s_j^{\text{Unit}}$$

and where $\hat{p}_j = p_j^{sale}\frac{A_i^{grid}\left(N_i, \{e_{it}^\star\}_{t=0}^T, \{A_{it}\}_{t=0}^T\right)}{A_i} + p_j\frac{A_i^{home}\left(N_i, \{e_{it}^\star\}_{t=0}^T, \{A_{it}\}_{t=0}^T\right)}{A_i}$ is the average of the purchasing and sales price of electricity, weighted by the fractions of electricity the household uses at home and sells back to the grid at the optimum.

For data on p_j^{sale}, we use the marginal cost of electricity as measured by Borenstein and Bushnell (2022). One challenge empirically is that we do not have dissagregated data on $A_i^{grid}(\cdot)$ or $A_i^{home}(\cdot)$. Therefore, we assume that the amount of electricity that is sold back to the grid is given by the reduced-form expression $A_i^{grid} = \left(N_i, \{e_{it}^\star\}_{t=0}^T, \{A_{it}\}_{t=0}^T\right) = A_i C N_i$, where C is a constant.

The households optimal number of panels is then given by

$$N_i^\star = \min\left[\bar{N}_i, -\left(\frac{\frac{\partial \mu_{ij}}{\partial N_i} + \gamma_{1N}}{2\left(\gamma_{2N} - C A_i\left(p_j - p_j^{sale}\right)\right)}\right)\right].$$

where, as before,

$$\mu_{ij}(N_i) = N_i A_i\left(p_j + s_j^{\text{kWh}}\right) - \left(1 - s_j^{\text{Cost}}\right) p_j^{\text{Ins}}(N_i) + N_i s_j^{\text{Unit}}.$$

We calibrate C such that a household with the average number of panels sells 30% of their electricity back to the grid.[75]

B.2 Maximum Likelihood Estimation of Power Plant Policy Functions

Let y^o_{kt} denote observed production from power plant k in time t, and let $\hat{f}_k\left(\text{ELoad}_t|\psi_k\right) = f_k\left(\text{ELoad}_t, \varepsilon_{kt}\right) - \varepsilon_{kt}$ denote the deterministic portion of the latent variable for power plant k in time t, written as a function the ψ^0_k, ψ^1_{Rk} and ψ^2_{Rk} parameters, which we collective denote by ψ_k. The log-likelihood contribution of a given hour of power plant k's production is

$$\begin{aligned}\log \mathcal{L}_{kt}\left(\text{ELoad}_t|\psi_k, \sigma^2_k\right) = & \mathbb{1}\left(y^o_{kt} = 0\right) \times \log\left(\Phi\left(\frac{\hat{f}_k\left(\text{ELoad}_t|\psi_k\right)}{\sigma_k}\right)\right) + \\ & \mathbb{1}\left(y^o_{kt} \in (0, \bar{y}_k)\right) \times \log\left(\frac{1}{\sigma_k}\phi\left(\frac{y^o_{kt} - \hat{f}_k\left(\text{ELoad}_t|\psi_k\right)}{\sigma_k}\right)\right) + \\ & \mathbb{1}\left(y^o_{kt} \geq \bar{y}_k\right) \times \log\left(1 - \Phi\left(\frac{\bar{y}_k - \hat{f}_k\left(\text{ELoad}_t|\psi_k\right)}{\sigma_k}\right)\right),\end{aligned} \tag{14}$$

where $\mathbb{1}(\cdot)$ represents an indicator functions which turns on if y^o_{kt} is equal to a given value or falls within a certain range, Φ is the standard normal CDF, and ϕ is the standard normal PDF. We choose the structural parameters for each power plant k by maximizing the sum of log likelihood contributions over all hours for that power plant. We restrict the parameter estimates such that output is weakly increasing in excess load for each region over the range of excess load observed in the data.

B.3 Details: Cost-Neutral Reforms

We can express the government's constrained maximization problem as the Lagrangian

$$W = \int_i V_i h(i) di - D\left(\text{ELoad}\right) - \lambda\left(\sum_j \int_{i \in I_j} s_{ij} m^\star_i h(i) di - G\right), \tag{15}$$

where, as before, $h(i)$ is the measure of households type i, $D\left(\text{ELoad}\right)$ is total environmental damages, $s_{ij} = s^{\text{Unit}}_j N^\star_i + s^{\text{kWh}}_j A_i N^\star_i + s^{\text{Cost}}_j p^{\text{Ins}}_j\left(N^\star_i\right)$ is the total subsidy paid to household i conditional on installation, and G is the maximum amount the government can spend on subsidies.

[75] https://www.seia.org/initiatives/net-metering

The optimal set of subsidies must satisfy the first-order conditions of the government's problem. Taking the derivative of W with respect to a given subsidy type $\theta \in \{\text{kWh}, \text{Unit}, \text{Cost}\}$ in state j yields

$$\frac{\partial W}{\partial s_j^\theta} = \int_i \frac{\partial V_i}{\partial s_j^\theta} h(i)di + \int_i \sum_{t=0}^{T} \frac{A_{it}}{(1+r)^t} \left| \frac{\partial D_t \left(\text{ELoad}_t^{SB}\right)}{\partial E_{Rt}^{\text{Solar}}} \right| \left(\overrightarrow{m_i}^\theta N_i^\star + m_i^\star \frac{\partial N_i}{\partial s_j^\theta} \right) h(i)di$$
$$- \lambda \left(\int_i \overrightarrow{m_i}^\theta s_{ij} h(i)di - \int_i m_i^\star \frac{\partial N_i}{\partial s_j^\theta} \frac{\partial s_{ij}}{\partial N_i} h(i)di - \int_i m_i^\star \frac{\partial s_{ij}}{\partial s_j^\theta} h(i)di \right), \quad (16)$$

where ELoad_t^{SB} denotes the excess load in time t evaluated at the optimal (welfare-maximizing) system of subsidies.

By the envelope theorem we have $\frac{\partial V_i}{\partial s_j^\theta} = m_i^\star \frac{\partial s_{ij}}{\partial s_j^\theta}$, which tells us that the utility gain for households is simply equal to the value of the increase in subsidy for inframarginal households, holding the number and size of installations constant.

Plugging this into (16) and setting the derivative equal to 0 yields

$$\int_i \sum_{t=0}^{T} \frac{A_{it}}{(1+r)^t} \left| \frac{\partial D_t \left(\text{ELoad}_t^{SB}\right)}{\partial E_{Rt}^{\text{Solar}}} \right| \left(\overrightarrow{m_i}^\theta N_i^\star + m_i^\star \frac{\partial N_i}{\partial s_j^\theta} \right) h(i)di$$
$$- \lambda \left(\int_i \overrightarrow{m_i}^\theta s_{ij} h(i)di - \int_i m_i^\star \frac{\partial N_i}{\partial s_j^\theta} \frac{\partial s_{ij}}{\partial N_i} h(i)di \right) - (\lambda - 1) \int_i m_i^\star \frac{\partial s_{ij}}{\partial s_j^\theta} h(i)di = 0, \quad (17)$$

This can be rewritten as

$$\int_i \overrightarrow{m_i}^\theta h(i)di \times \left(\frac{\int_i \overrightarrow{m_i}^\theta \sum_{t=0}^{T} \frac{A_{it}}{(1+r)^t} \left| \frac{\partial D_t \left(\text{ELoad}_t^{SB}\right)}{\partial E_{Rt}^{\text{Solar}}} \right| N_i^\star}{\int_i \overrightarrow{m_i}^\theta h(i)di} - \lambda \frac{\int_i \overrightarrow{m_i}^\theta s_{ij} h(i)di}{\int_i \overrightarrow{m_i}^\theta h(i)di} \right) +$$
$$\int_i \frac{\partial N_i}{\partial s_j^\theta} h(i) \times \left(\frac{\int_i m_i^\star \frac{\partial N_i}{\partial s_j^\theta} \sum_{t=0}^{T} \frac{A_{it}}{(1+r)^t} \left| \frac{\partial D_t \left(\text{ELoad}_t^{SB}\right)}{\partial E_{Rt}^{\text{Solar}}} \right|}{\int_i \frac{\partial N_i}{\partial s_j^\theta} h(i)} - \lambda \frac{\int_i \frac{\partial N_i}{\partial s_j^\theta} \frac{\partial s_{ij}}{\partial N_i} h(i)}{\int_i \frac{\partial N_i}{\partial s_j^\theta} h(i)} \right) +$$
$$(1 - \lambda) M_j \frac{\int_i m_i^\star \frac{\partial s_{ij}}{\partial s_j^\theta} h(i)di}{\int_i m_i^\star h(i)di}. \quad (18)$$

Finally, plugging in the definitions of $\frac{\partial M_j}{\partial s_j^\theta}$, $\overrightarrow{\Delta D}_j^{\theta,\text{ext}}$, $\overrightarrow{s}_j^{\theta,\text{ext}}$, $\left.\frac{\partial N_j}{\partial s_j^\theta}\right|_{M_j^{\text{st}}}$, $\overrightarrow{\Delta D}_j^{\theta,\text{int}}$, $\overrightarrow{\frac{\partial s}{\partial N}}_j^{\theta,\text{int}}$, and $\overline{\frac{\partial s_{ij}}{\partial s_j^\theta}}$ yields (8).

B.4 Numerical Algorithm for Calculating Optimal Subsidies

In this appendix, we outline the numerical algorithm we use to solve for the welfare-maximizing subsidies.

1. Make a guess of the marginal value of public funds, λ. Call this guess $\hat{\lambda}$.

2. Make a guess of the set of subsidies. Let this matrix of all types of subsidies in all states be denoted by $\hat{s}$.

3. Given the current guess of subsidies, $\hat{s}$, and the guess of the marginal value of public funds, $\hat{\lambda}$, calculate the first-order conditions of the government's problem for each subsidy type and each state given by (8). We use analytical derivatives to evaluate $\frac{\partial M_j}{\partial s_j^\theta}$, $\overrightarrow{\Delta D}_j^{\theta,\text{ext}}$, $\overrightarrow{s}_j^{\theta,\text{ext}}$, $\left.\frac{\partial N_j}{\partial s_j^\theta}\right|_{M_j^{\text{st}}}$ $\overrightarrow{\Delta D}_j^{\theta,\text{int}}$, and $\overrightarrow{\frac{\partial s}{\partial N}}_j^{\theta,\text{int}}$.

4. If all of the first-order conditions are sufficiently close to 0, move on to the next step. If not, update the guess of the subsidies are return to Step 3.

5. Given the current guess of subsidies, calculate the total government cost.

6. If the government cost is sufficiently close to G, then the current guesses, $\hat{\lambda}$ and $\hat{s}$, solve the constrained maximization problem. If not, take a new guess for $\hat{\lambda}$ and return to Step 2.

B.5 Details: Damage-Minimizing Subsidies

The government's problem is to choose subsidies to minimize national damages, $D\,(\text{ELoad})$, subject to the budget constraint that the total spending on subsidies cannot exceed some value G:

$$\sum_j \int_{i \in I_j} s_{ij} m_i^\star h(i) di \leq G,$$

where $h(i)$ is the measure of households type i, $s_{ij} = s_j^{\text{Unit}} N_i^\star + s_j^{\text{kWh}} A_i N_i^\star + s_j^{\text{Cost}} p_j^{\text{Ins}} (N_i^\star)$ is the total subsidy paid to household i conditional on installation, and G is the maximum amount the government can spend on subsidies.

We can express this constrained optimization problem as the Lagrangian

$$W = -D\,(\text{ELoad}) - \lambda \left(\sum_j \int_{i \in I_j} s_{ij} m_i^\star h(i) di - G \right). \tag{19}$$

Taking the derivative of W with respect to s_j^θ yields

$$\frac{\partial W}{\partial s_j^\theta} = \int_i \sum_{t=0}^{T} \frac{A_{it}}{(1+r)^t} \left| \frac{\partial D_t\left(\mathrm{ELoad}_t^{MD}\right)}{\partial E_{Rt}^{\mathrm{Solar}}} \right| \left(\overrightarrow{m_i}^\theta N_i^\star + m_i^\star \frac{\partial N_i}{\partial s_j^\theta} \right) h(i) di$$
$$- \lambda \left(\int_i \overrightarrow{m_i}^\theta s_{ij} h(i) di - \int_i m_i^\star \frac{\partial N_i}{\partial s_j^\theta} \frac{\partial s_{ij}}{\partial N_i} h(i) di - \int_i m_i^\star \frac{\partial s_{ij}}{\partial s_j^\theta} h(i) di \right), \quad (20)$$

where ELoad_t^{MD} denotes the excess load in time t evaluated at the optimal (damage-minimizing) system of subsidies.

As in Appendix B.3, we can again use $\frac{\partial V_i}{\partial s_j^\theta} = m_i^\star \frac{\partial s_{ij}}{\partial s_j^\theta}$ by the envelope theorem. Plugging this in and rearranging, we can write the first-order condition as

$$\int_i \overrightarrow{m_i}^\theta h(i) di \times \left(\frac{\int_i \overrightarrow{m_i}^\theta \sum_{t=0}^{T} \frac{A_{it}}{(1+r)^t} \left| \frac{\partial D_t\left(\mathrm{ELoad}_t^{MD}\right)}{\partial E_{Rt}^{\mathrm{Solar}}} \right| N_i^\star}{\int_i \overrightarrow{m_i}^\theta h(i) di} - \lambda \frac{\int_i \overrightarrow{m_i}^\theta s_{ij} h(i) di}{\int_i \overrightarrow{m_i}^\theta h(i) di} \right) +$$
$$\int_i \frac{\partial N_i}{\partial s_j^\theta} h(i) \times \left(\frac{\int_i m_i^\star \frac{\partial N_i}{\partial s_j^\theta} \sum_{t=0}^{T} \frac{A_{it}}{(1+r)^t} \left| \frac{\partial D_t\left(\mathrm{ELoad}_t^{MD}\right)}{\partial E_{Rt}^{\mathrm{Solar}}} \right|}{\int_i \frac{\partial N_i}{\partial s_j^\theta} h(i)} - \lambda \frac{\int_i \frac{\partial N_i}{\partial s_j^\theta} \frac{\partial s_{ij}}{\partial N_i} h(i)}{\int_i \frac{\partial N_i}{\partial s_j^\theta} h(i)} \right) +$$
$$- \lambda M_j \frac{\int_i m_i^\star \frac{\partial s_{ij}}{\partial s_j^\theta} h(i) di}{\int_i m_i^\star h(i) di}. \quad (21)$$

Finally, plugging in the definitions of $\frac{\partial M_j}{\partial s_j^\theta}$, $\overrightarrow{\Delta D}_j^{\theta,\mathrm{ext}}$, $\overrightarrow{s}_j^{\theta,\mathrm{ext}}$, $\left.\frac{\partial N_j}{\partial s_j^\theta}\right|_{M_j^{\mathrm{st}}}$ $\overrightarrow{\Delta D}_j^{\theta,\mathrm{int}}$, $\overrightarrow{\frac{\partial s}{\partial N}}_j^{\theta,\mathrm{int}}$, and $\overline{\frac{\partial s_{ij}}{\partial s_j^\theta}}$ yields (22), which gives the first-order condition for each subsidy type in each state j:

$$\underbrace{\frac{\partial M_j}{\partial s_j^\theta} \times \left(\overrightarrow{\Delta D}_j^{\theta,\mathrm{ext}} - \lambda \overrightarrow{s}_j^{\theta,\mathrm{ext}} \right)}_{\text{Extensive Margin}} + \underbrace{\left.\frac{\partial N_j}{\partial s_j^\theta}\right|_{M_j^{\mathrm{st}}} \times \left(\overrightarrow{\Delta D}_j^{\theta,\mathrm{int}} - \lambda \overrightarrow{\frac{\partial s}{\partial N}}_j^{\theta,\mathrm{int}} \right)}_{\text{Intensive Margin}} - \underbrace{\lambda M_j \overline{\frac{\partial s_{ij}}{\partial s_j^\theta}}}_{\text{Mechanical Effect}} = 0. \quad (22)$$

These optimality conditions for a damage-minimizing planner share a similar structure to those of the welfare-maximizing planner given by (8). The exception is how the two planners value increases in subsidies given to inframarginal households, which are represented the third term in each of the first-order conditions ("Mechanical Effect"). For the damage-minimizing planner, increases subsidies for these inframarginal households entail a fiscal cost with no additional decrease in damages. Therefore, the number of inframarginal households (M_j)

enters negatively into the first order condition. The welfare-maximizing planner, on the other hand, values the increase in utility associated with increases in subsidies for inframarginal households. Therefore, each additional dollar of subsidies for an inframarginal household is valued at $(1-\lambda)$, reflecting both this increase in utility and the fiscal cost.

B.6 Details: Net-Cost Neutral Reforms

In this section we consider a planner who maximizes the sum of utility subject to a net-cost budget constraint, where environmental damages are counted as a fiscal cost. Specifically, the government maximizes $\int_i V_i h(i) di$ subject to the constraint

$$D\left(\text{ELoad}\right) + \sum_j \int_{i \in I_j} s_{ij} m_i^\star h(i) di \leq \tilde{G}, \tag{23}$$

where $\tilde{G}$ is the maximum net-cost the government can take on.

We can again express the government's problem as a Lagrangian

$$W = \int_i V_i h(i) di - \lambda \left(D\left(\text{ELoad}\right) + \sum_j \int_{i \in I_j} s_{ij} m_i^\star h(i) di - \tilde{G} \right), \tag{24}$$

where $h(i)$ is the measure of households type i, $D\left(\text{ELoad}\right)$ is total environmental damages, and $s_{ij} = s_j^{\text{Unit}} N_i^\star + s_j^{\text{kWh}} A_i N_i^\star + s_j^{\text{Cost}} p_j^{\text{Ins}}\left(N_i^\star\right)$ is the total subsidy paid to household i conditional on installation.

Taking the derivative of W with respect to a given subsidy type $\theta \in \{\text{kWh}, \text{Unit}, \text{Cost}\}$ in state j yields and following the logic from Appendix B.3, we arrive at the first-order conditions which characterize the optimal set of subsidies:

$$\underbrace{\frac{\partial M_j}{\partial s_j^\theta} \times \left(\lambda \overrightarrow{\Delta D}_j^{\theta,\text{ext}} - \lambda \overrightarrow{s}_j^{\theta,\text{ext}} \right)}_{\text{Extensive Margin}} + \underbrace{\left.\frac{\partial N_j}{\partial s_j^\theta}\right|_{M_j^{\text{st}}} \times \left(\lambda \overrightarrow{\Delta D}_j^{\theta,\text{int}} - \lambda \overrightarrow{\frac{\partial s}{\partial N}}_j^{\theta,\text{int}} \right)}_{\text{Intensive Margin}} - \underbrace{\left(1-\lambda\right) M_j \overline{\frac{\partial s_{ij}}{\partial s_j^\theta}}}_{\text{Mechanical Effect}} = 0. \tag{25}$$

Results The results are summarized in Table A3. The results are fairly similar to the welfare-maximizing reforms that we present as our main results; we find that switching to the optimal net-cost neutral reform leads to over a 7% increase in aggregate environmental benefits of solar panels.

	(1)	(2)
	Baseline	Net Cost Neutral
I. Average Subsidy ($Thousands)		
Midwest	11.3	16.2
Northeast	18.2	17.4
South	10.4	15.3
West	12.0	13.9
II. Installations per 1000HHs		
Midwest	4.1	6.3
Northeast	29.8	17.5
South	6.8	11.2
West	13.2	16.4
National	11.4	12.6
III. Annual Damages Offset ($Millions)		
CO2e	69.8	76.2
NOx	18.0	19.5
PM2.5	16.7	16.3
SO2	37.0	35.1
Total	141.5	147.1

Table A3: Net-cost neutral reforms. The first panel shows the average present discounted value of subsidies received for a 15-panel installation for each census region. The second panel gives the simulated number of solar installations per 1000 households in the model for each Census region. The final panel gives the total damages offset by rooftop solar.

C Results Appendix

C.1 Installation Prices

Table A4 shows the results for estimating solar system installation prices using the Tracking the Sun data using the following regression,

$$p^{\text{Ins}}_{R(j)}(N_i) = p^{0,\text{Ins}}_{R(j)} + p^{1,\text{Ins}}_{R(j)} N_i + \varepsilon_{ij}, \tag{26}$$

where $p^{0,\text{Ins}}_{R(j)}$ is a fixed cost and $p^{1,\text{Ins}}_{R(j)}$ is a per-panel cost, and $R(j)$ is the Census Region containing state j. The table shows results for the full sample and each region, where the intercept gives the fixed installation cost, and the coefficient on the number of panels is the per-panel cost. The linear model is a good fit for the data, as seen in Figure A7, which shows our fitted line against a flexible smoothing function for each region.

Dep. Var.:	Total Cost				
Census Region	Full sample	Midwest	Northeast	South	West
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
(Intercept)	5,960.9***	10,129.9***	3,558.1***	7,772.5***	6,295.5***
	(28.8)	(610.0)	(57.0)	(283.1)	(33.9)
Num. Panels	1,078.7***	1,051.7***	1,156.7***	823.6***	1,071.5***
	(1.4)	(27.7)	(2.4)	(11.9)	(1.8)
Fit statistics					
Observations	1,273,431	1,097	254,336	22,245	995,753
R^2	0.55	0.53	0.67	0.54	0.51
Adjusted R^2	0.55	0.53	0.67	0.54	0.51

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table A4: Solar system installation prices.

C.2 Relationship Between Installations and Monetary Incentives

Figure A8 plots the average monetary benefit associated with installing a 15-panel system in each state against the log installations per capita in each state.

Table A5 regresses tract-level log installations on the monetary benefits of installation, where again we calculate the monetary benefits of installation as $\mu_{ij}(N_i^\star)$ evaluated at

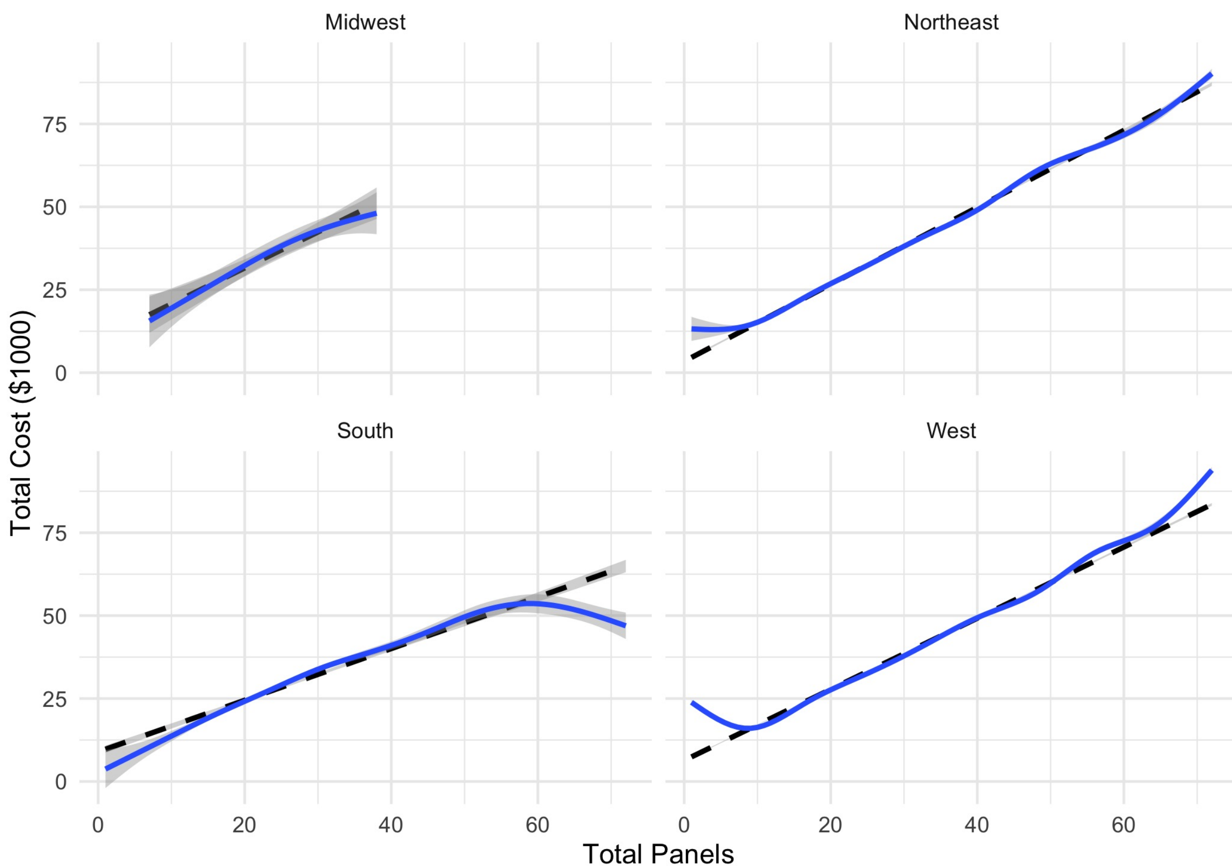


Figure A7: Estimation results for solar system price regression, where the dashed black line is our estimated model and the solid blue line shows the fit of a generalized additive model.

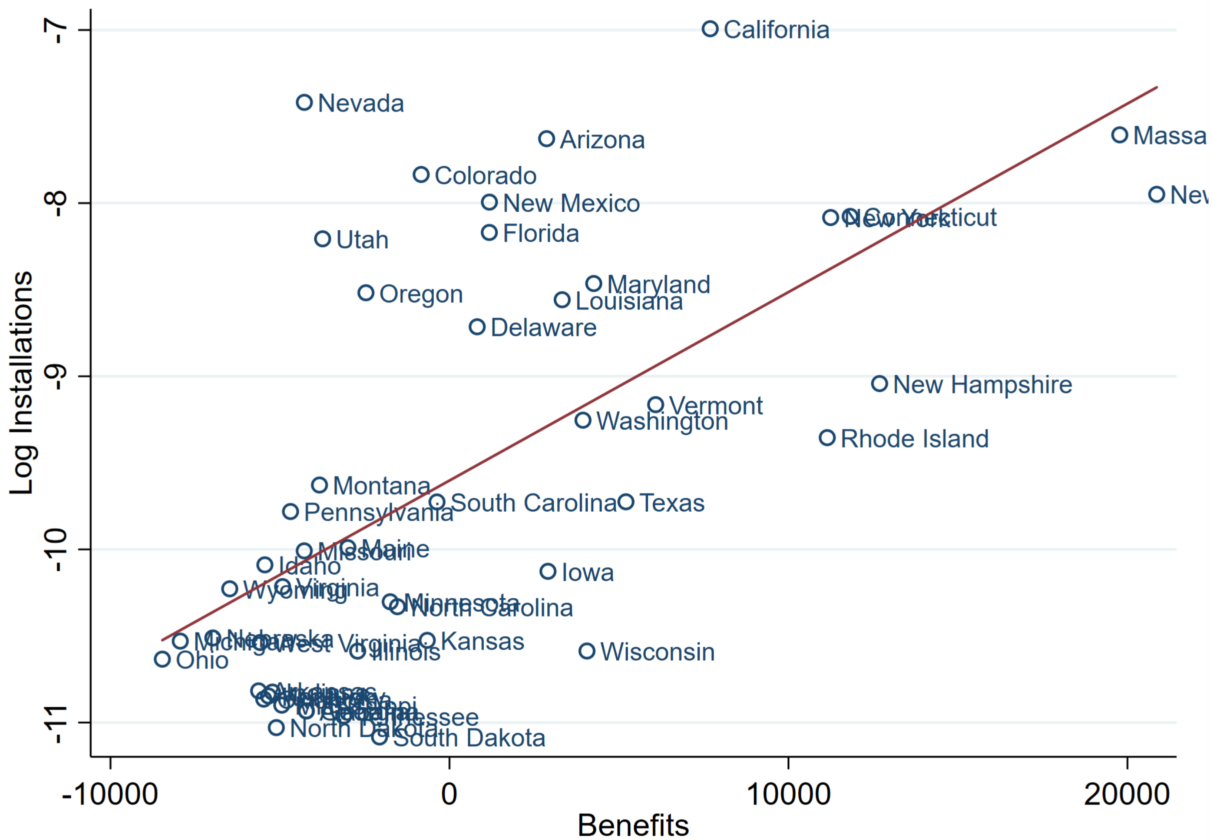


Figure A8: Monetary benefits and log installations per capita by state. The Y-axis plots plots the average monetary benefit associated with installing a 15-panel system in each state. The X-axis gives log installations per capita in each state.

	(1)	(2)	(3)	(4)	(5)	(6)
Financial Benefits	0.105***	0.104***	0.0793***	0.0820***	0.0865***	0.0892***
	(0.0332)	(0.0330)	(0.0158)	(0.0186)	(0.0157)	(0.0176)
Observations	41,776	41,776	41,776	41,776	41,776	41,776
R-squared	0.187	0.201	0.388	0.411	0.421	0.445
Demographic Controls	NO	YES	NO	YES	NO	YES
Region FE	NO	NO	YES	YES	NO	NO
Division FE	NO	NO	NO	NO	YES	YES

*** p<0.01, ** p<0.05, * p<0.1

Table A5: Regression of log installations on the net present value of total financial benefits associated with solar panel installations. Financial benefits measured in thousands of dollars. Standard errors clustered by state.

	(1)	(2)	(3)	(4)	(5)	(6)
Financial Benefits	0.0812***	0.0920***	0.0770***	0.0855***	0.108***	0.111***
	(0.0226)	(0.0215)	(0.0224)	(0.0220)	(0.0308)	(0.0313)
Observations	41,776	41,776	41,776	41,776	41,776	41,776
R-squared	0.013	0.018	0.027	0.030	0.042	0.043
Demographic Controls	NO	YES	NO	YES	NO	YES
Region FE	NO	NO	YES	YES	NO	NO
Division FE	NO	NO	NO	NO	YES	YES

*** p<0.01, ** p<0.05, * p<0.1

Table A6: Regression of average panel size on the net present value of total financial benefits associated with solar panel installations. Financial benefits measured in thousands of dollars. Standard errors clustered by state.

$N_i^\star = 15$, the average number of panels in a solar system in the data. Specifications with "Demographic Controls" include controls for tract-level college completion percentage and percent democrat. Columns (3) and (4) add Census region fixed effects while columns (5) and (6) include Census division fixed effects. Across all specifications, we find that a \$1000 increase in monetary benefits for a 15-panel installation is associated with a 7.9% to 10.5% increase in installations.

C.3 Additional Regressions

Table A6 regresses tract-level data on average number of panels per installation on monetary benefits of installation, where again we calculate the monetary benefits of installation as $\mu_{ij}\left(N_i^\star\right)$ evaluated at $N_i^\star = 15$, the average number of panels in a solar system in the data. Across all specifications, we find that a \$1000 increase in monetary benefits is associated with a 0.08 to 0.11 increase in average panels per installation.

	SD Installs
Baseline	1
Harmonize Subsidies	0.48
Harmonize Electricity Prices	0.25
Harmonize Installation Prices	0.23
Equalize Sunlight	0.09
Harmonize Demographics	0.08

Table A7: Each row shows the standard deviation in installation rates across states relative to the baseline level. Each row equalizes a given object across tracts and recalculates the standard deviation across states. In each simulation, we set the object in question at a level such that the total national installations remains at the baseline value.

C.4 Decomposition of Differences in Installation Rates

Solar panel installation rates may differ across states for five main reasons: 1) subsidies, 2) electricity prices, 3) installation prices, 4) sunlight, and 5) household demographics. We sequentially equalize each of these five factors across states and re-simulate the model.

First, we examine the role of subsidies by harmonizing subsidies across states. Specifically, we set all subsidies equal to the population-weighted average across states. We then multiply these subsidies by a constant such that the total number of installations nationally in this counterfactual environment equals those in the baseline simulation. Table A7 shows the standard deviation in installation rates across states relative to the standard deviation in the baseline. Equalizing subsidies across states leads to a 52% drop in the standard deviation of installation rates, suggesting that variation in state solar subsidies explains a large proportion of the variation in installation rate across states.[76]

The following rows sequentially equalize electricity prices, installation prices, solar irradiance, and demographics. In each counterfactual, we set the object in question at a level such that the total national installations equal the baseline value. The results show that electricity prices and solar irradiance play the most important roles in explaining the remaining variation in installation rates, with installation prices only playing a minor role.[77] Taken together, these results suggest that differences in subsidies and electricity prices across states are the most important drivers of differences in installation rates.

Table A8 repeats the analysis from above but equalizes electricity prices before equalizing subsidy levels across states. Again, we can see that electricity prices and subsidies explain the majority of the deviation in installation rates across states.

[76] In a second decomposition below, we simulate first equalizing energy prices before harmonizing subsidies. We again find that energy subsidies and electricity prices are the two most important factors explaining differences in installation rates across states.

[77] The remaining variation is due to differences in the distribution of building size ($\bar{N}_i$) across states.

	SD Installs
Baseline	1
Harmonize Electricity Prices	0.79
Harmonize Subsidies	0.25

Table A8: Alternative decomposition. Each row shows the standard deviation in installation rates across states relative to the baseline level. Each row equalizes a given object across states and recalculates the standard deviation across states. In each simulation, we set the object in question at a level such that the total national installations remains at the baseline value.

C.5 Power Plant Model

Here we present additional information about the power plant model estimation results. Figures A9 and A10 show the model fit graphs for the main paper, broken out into regions. These show that model performance is consistent within each interconnection. A11 shows how fuel mix varies for each region.

Additionally, the model does a good job predicting total production from power plants. Figure A12 shows the average hourly production for each fossil-fuel plant, comparing the actual data to our estimated data. The predicted values are very close to the actual values throughout the range of actual power plant production, indicating that the model can predict individual plants' production very well on average.

Figure A13 shows a histogram of estimated marginal damages for one kWh of production. These are broken up by stack height and whether the plant's production is above or below its median value.

C.6 Average Panel Size Across Counterfactuals

Table A9 shows the average number of panels per installation across Census regions in each simulation. We can see that panel size does not significantly change across regions or across simulations. These results suggest that extensive-margin adjustments shown in the body of the paper play a much more important role quantitatively than the intensive-margin adjustments shown here.

C.7 State-Level Results

The first columns of Table A10 gives the baseline and welfare-maximizing subsidy in each state. The following columns show the simulated number of solar panel installations per 1000 households under the current subsidies and under the welfare-maximizing subsidies.

The first two columns of Table A11 shows the state-level subsidies given the current system and the damage-minimizing subsidies. The following columns of Table A11 show the

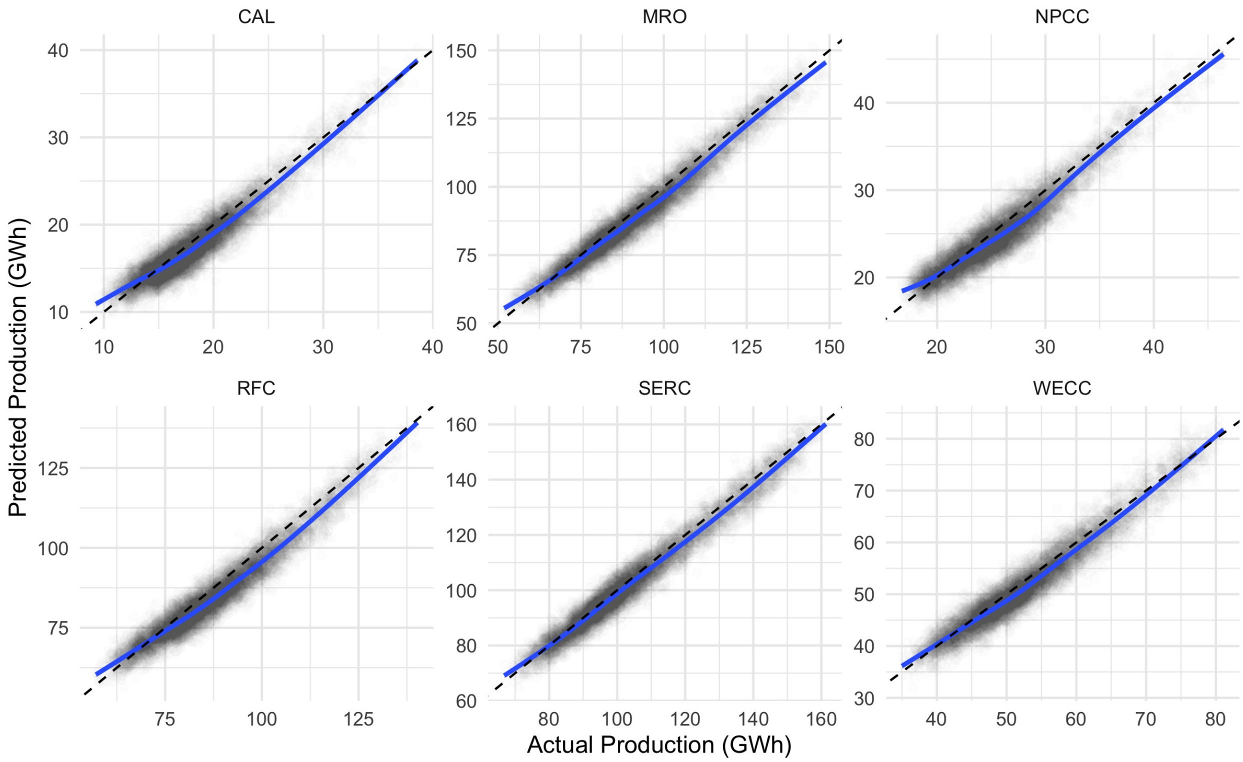


Figure A9: Model fit at the region level, excluding Texas as there is only one region in the Texas interconnection. Dots represent an hour of production for each region, smoothed lines show the fit of a generalized additive model.

	(1)	(2)	(3)	(4)	(5)
		State-Specific Subsidies		Tract-Specific Subsidies	
	Baseline	Welfare Max	Damage Min	Welfare Max	Damage Min
I. Average Number of Panels per installation					
Midwest	14.82	14.87	14.88	14.87	14.89
Northeast	14.89	14.89	14.90	14.89	14.90
South	14.82	14.88	14.88	14.88	14.88
West	14.85	14.88	14.84	14.88	14.84

Table A9: Each entry gives the average number of solar panels in a solar installation across Census regions in each model simulation.

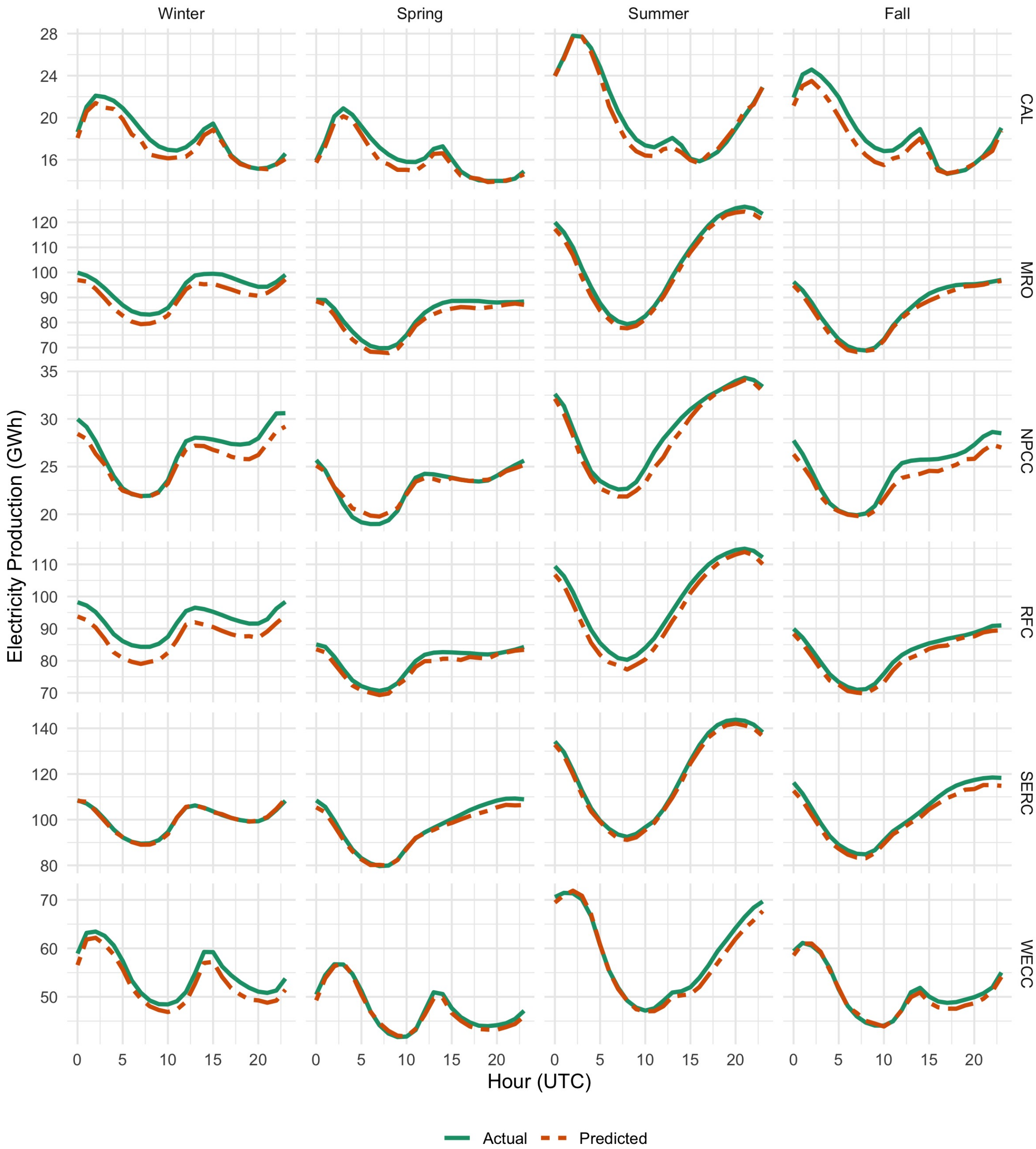


Figure A10: Model fit at the region level by hour and season.

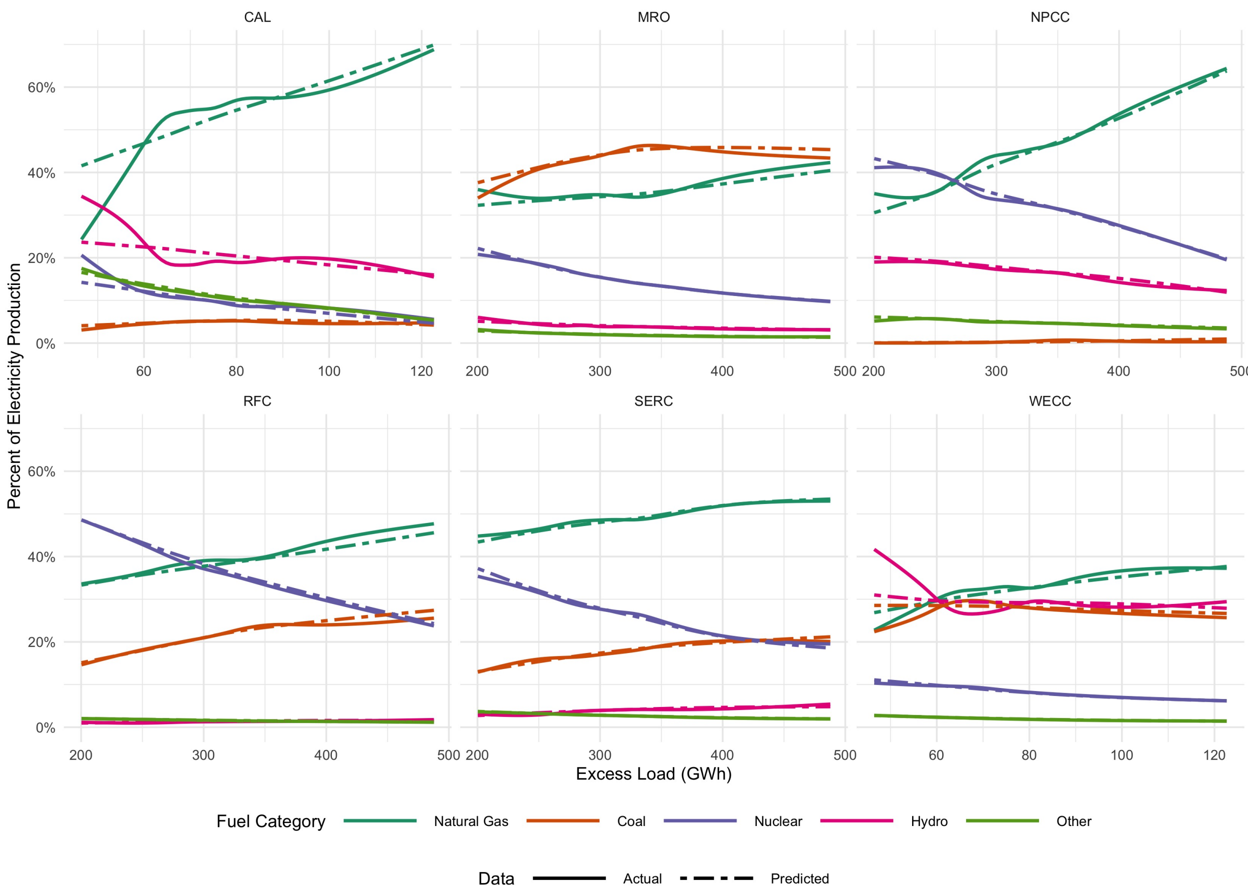


Figure A11: Fuel mix of production by reigon. The X-axis gives excess load at the interconnection level and the Y-axis gives the percent of electricity production that is produced by each of the fuel types. The dashed lines show the fuel mix in the data while the solid lines show the simulated fuel mix.

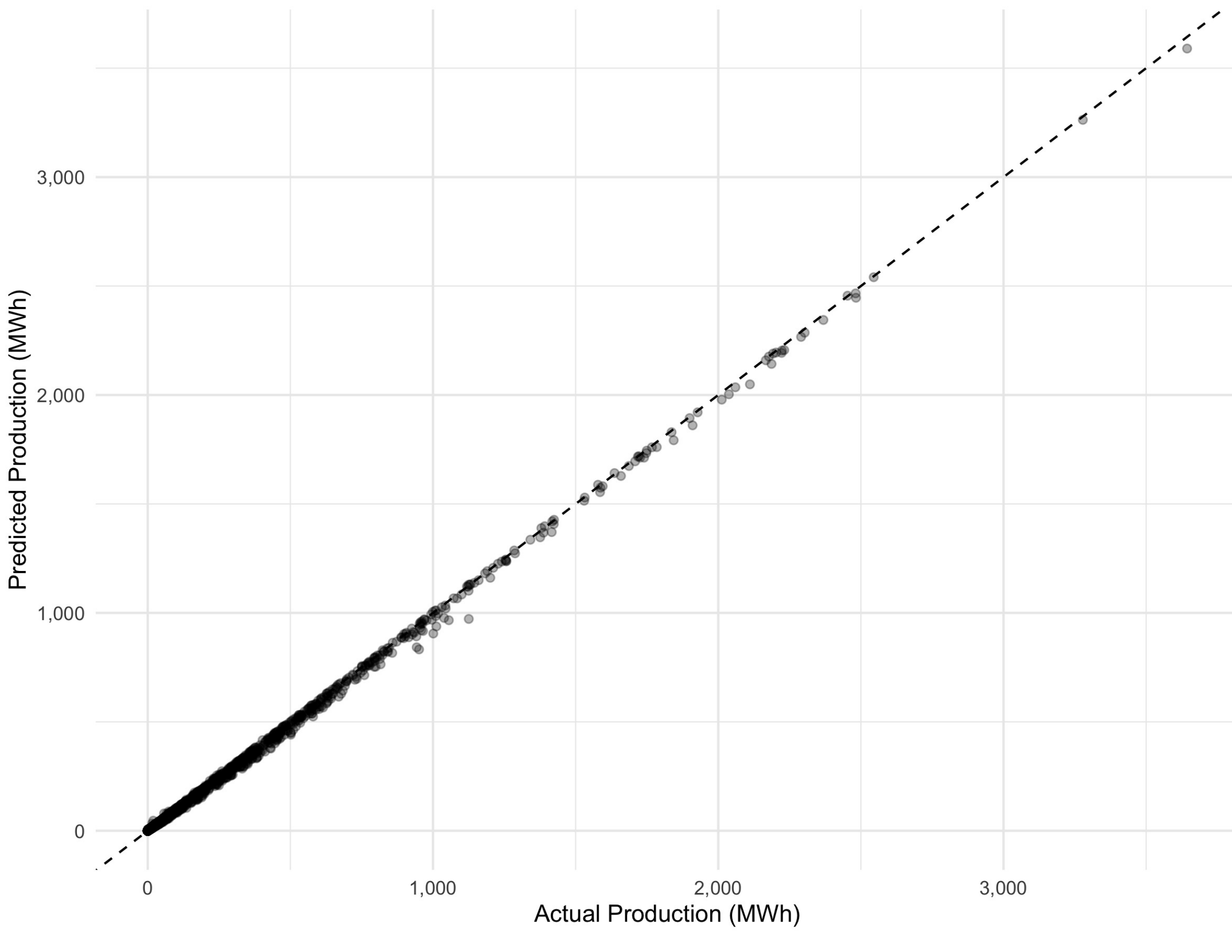


Figure A12: Model fit of average production for each power plant. Each point represents one of the 4,625 dispatchable power plants that we estimate parameters for in the model.

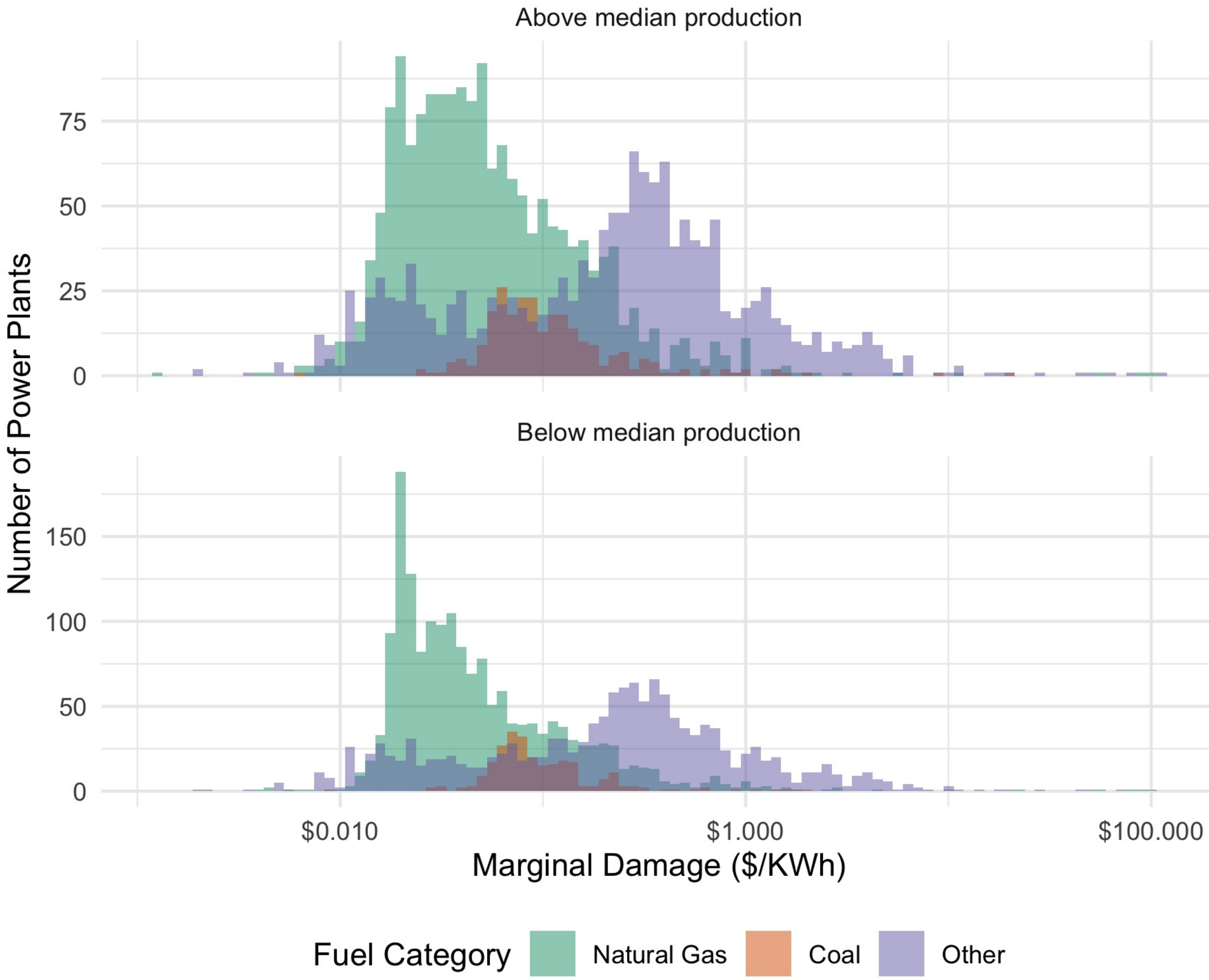


Figure A13: Marginal damages by power plant.

	Expected Subsidy		Installs/1000HH			Expected Subsidy		Installs/1000HH	
	Baseline	Optimal	Baseline	Optimal		Baseline	Optimal	Baseline	Optimal
Alabama	5.6	15.2	2.8	8.5	Nebraska	8.8	14.6	2.7	5.3
Arizona	11.5	13.0	9.2	10.9	Nevada	6.8	12.9	4.0	8.1
Arkansas	5.6	14.7	2.9	8.1	New Hampshire	19.2	18.0	22.7	19.5
California	12.0	12.7	18.0	19.7	New Jersey	29.0	19.5	64.1	22.6
Colorado	9.8	12.3	5.7	7.6	New Mexico	11.0	13.2	8.2	10.5
Connecticut	16.9	18.1	23.9	27.4	New York	18.6	18.1	21.1	20.0
Delaware	9.8	19.8	7.0	21.6	North Carolina	8.6	15.2	5.4	11.4
Florida	10.0	15.7	7.6	14.5	North Dakota	11.6	13.9	2.9	3.7
Georgia	5.6	15.2	4.1	12.5	Ohio	7.5	18.9	2.2	8.0
Idaho	7.7	11.7	2.8	4.5	Oklahoma	5.6	15.1	3.0	8.9
Illinois	13.5	18.5	4.9	8.7	Oregon	11.5	10.9	5.3	4.9
Indiana	11.1	19.3	3.2	8.1	Pennsylvania	6.2	19.2	3.1	13.6
Iowa	19.6	14.2	8.1	4.3	Rhode Island	17.0	18.2	20.6	23.6
Kansas	13.6	14.9	5.7	6.5	South Carolina	9.1	15.3	5.2	10.9
Kentucky	6.1	14.8	3.1	8.5	South Dakota	13.1	13.4	3.1	3.2
Louisiana	15.0	14.9	9.2	9.1	Tennessee	7.4	14.9	3.5	8.2
Maine	5.8	18.0	3.7	14.9	Texas	15.7	13.3	9.9	7.6
Maryland	12.2	19.9	10.8	25.6	Utah	8.2	12.1	3.3	5.2
Massachusetts	25.6	18.1	56.9	24.9	Vermont	14.0	17.6	15.6	23.6
Michigan	7.2	13.7	2.3	4.8	Virginia	5.6	15.8	3.5	11.2
Minnesota	14.0	13.9	5.1	5.0	Washington	19.1	10.7	10.7	4.1
Mississippi	5.6	15.2	3.1	9.4	West Virginia	5.6	19.2	2.3	10.9
Missouri	10.9	14.6	3.8	5.7	Wisconsin	19.2	17.5	9.2	7.5
Montana	9.4	11.4	3.5	4.4	Wyoming	6.2	15.1	2.0	12.4

Table A10: The first two columns shows the state-level subsidies given the current system and the welfare-maximizing subsidies. Subsidies are measured as the average present discounted value of subsidies for a 15-panel installation, measured in thousands of dollars. The following columns show the simulated number of solar panel installations per 1000 households given the current subsidies and under the welfare-maximizing subsidies.

simulated number of solar panel installations per 1000 households given the current subsidies and under the damage-minimizing subsidies.

C.8 Marginal Subsidy Increases: Results

Table A12 shows the damages offset per additional dollar of government funds associated with marginal subsidy increases relative to the current system of subsidies. The first column ("Unit") gives the damages offset per dollar associated with marginal increases in unit-based subsidies, s_j^{Unit}, the second column ("Cost") gives the damages offset per dollar associated with marginal increases in cost-based subsidies, s_j^{Cost}, and the third column ("kWh") gives the damages offset per dollar associated with marginal increases in production-based subsidies, s_j^{kWh}. For example, the first entry in the table tells us that a small increase in unit-based subsidies in Alabama would lead to a 32-cent decrease in damages for each additional dollar of government funds.

The three states with the largest damages offset per dollar are West Virginia, Pennsylvania, and Maine. The three states with smallest damages offset are Washington, Oregon, and Montana. Within a state, there are only small differences in the cost-effectiveness of various subsidy types.

D Extensions and Robustness Appendix

D.1 Alternative Specifications of Household Utility

Table A13 recalculates our main results under alternative specifications of household utility. Each entry shows the change in average subsidies, installations, and environmental benefits associated with moving from the current system of subsidies to the welfare-maximizing system of subsidies, given the specification in question. Column (1) considers a specification in which the nonpecuniary component does not depend on tract-level demographics. In (2), we add the tract-level fraction of individuals with a college education. In (3), we also add the fraction of democrat voters, and in (4), we add the homeownership rate. Note that Column (3) is the same as our baseline specification. The results are qualitatively very similar across specifications.

D.2 Line Losses

We use the methodology from Borenstein and Bushnell (2022) to account for line losses between the power plant and households. Formally, losses for each region come from a constant

	Expected Subsidy		Installs/1000HH			Expected Subsidy		Installs/1000HH	
	Baseline	Optimal	Baseline	Optimal		Baseline	Optimal	Baseline	Optimal
Alabama	5.6	13.3	2.8	6.8	Nebraska	8.8	11.7	2.7	3.8
Arizona	11.5	7.4	9.2	5.7	Nevada	6.8	7.2	4.0	4.2
Arkansas	5.6	11.9	2.9	5.9	New Hampshire	19.2	20.0	22.7	24.6
California	12.0	6.3	18.0	9.4	New Jersey	29.0	23.8	64.1	36.7
Colorado	9.8	5.6	5.7	3.5	New Mexico	11.0	7.8	8.2	5.7
Connecticut	16.9	20.2	23.9	34.5	New York	18.6	20.5	21.1	26.0
Delaware	9.8	24.6	7.0	37.2	North Carolina	8.6	13.1	5.4	8.9
Florida	10.0	14.5	7.6	12.7	North Dakota	11.6	9.5	2.9	2.3
Georgia	5.6	13.1	4.1	9.8	Ohio	7.5	23.1	2.2	12.9
Idaho	7.7	4.4	2.8	1.9	Oklahoma	5.6	13.2	3.0	7.1
Illinois	13.5	21.9	4.9	12.8	Oregon	11.5	2.0	5.3	1.8
Indiana	11.1	24.1	3.2	14.1	Pennsylvania	6.2	23.4	3.1	22.0
Iowa	19.6	10.5	8.1	2.8	Rhode Island	17.0	20.5	20.6	30.5
Kansas	13.6	12.4	5.7	4.9	South Carolina	9.1	13.5	5.2	8.9
Kentucky	6.1	12.0	3.1	6.2	South Dakota	13.1	8.6	3.1	1.8
Louisiana	15.0	12.4	9.2	6.8	Tennessee	7.4	12.4	3.5	6.2
Maine	5.8	20.4	3.7	19.5	Texas	15.7	8.6	9.9	4.4
Maryland	12.2	24.7	10.8	43.4	Utah	8.2	5.2	3.3	2.3
Massachusetts	25.6	20.0	56.9	31.0	Vermont	14.0	19.0	15.6	27.6
Michigan	7.2	9.3	2.3	2.9	Virginia	5.6	14.6	3.5	9.7
Minnesota	14.0	9.8	5.1	3.1	Washington	19.1	1.5	10.7	1.4
Mississippi	5.6	13.2	3.1	7.4	West Virginia	5.6	23.7	2.3	18.2
Missouri	10.9	11.6	3.8	4.1	Wisconsin	19.2	19.3	9.2	9.2
Montana	9.4	3.1	3.5	1.7	Wyoming	6.2	4.9	2.0	1.7

Table A11: The first two columns shows the state-level subsidies given the current system and the damage-minimizing subsidies. Subsidies are measured as the average present discounted value of subsidies for a 15-panel installation, measured in thousands of dollars. The following columns show the simulated number of solar panel installations per 1000 households given the current subsidies and under the damage-minimizing subsidies.

	Subsidy Type				Subsidy Type		
	Unit	Cost	kWh		Unit	Cost	kWh
Alabama	32.1%	32.0%	32.1%	Nebraska	24.3%	24.3%	24.3%
Arizona	13.7%	13.7%	13.7%	Nevada	21.5%	21.5%	21.5%
Arkansas	30.1%	30.1%	30.1%	New Hampshire	21.5%	21.5%	21.5%
California	15.0%	15.0%	15.1%	New Jersey	17.9%	17.9%	17.9%
Colorado	16.1%	16.1%	16.1%	New Mexico	17.5%	17.5%	17.5%
Connecticut	23.7%	23.7%	23.7%	New York	22.3%	22.3%	22.4%
Delaware	37.9%	37.9%	37.9%	North Carolina	26.2%	26.2%	26.2%
Florida	26.0%	25.9%	26.0%	North Dakota	19.0%	19.0%	19.0%
Georgia	32.0%	31.9%	32.0%	Ohio	41.0%	41.0%	41.1%
Idaho	16.8%	16.7%	16.8%	Oklahoma	32.0%	32.0%	32.0%
Illinois	29.1%	29.1%	29.1%	Oregon	9.1%	9.1%	9.2%
Indiana	34.7%	34.7%	34.8%	Pennsylvania	45.3%	45.3%	45.4%
Iowa	14.2%	14.2%	14.2%	Rhode Island	23.8%	23.8%	23.8%
Kansas	19.7%	19.7%	19.7%	South Carolina	24.5%	24.5%	24.5%
Kentucky	16.9%	16.9%	16.9%	South Dakota	16.9%	16.9%	16.8%
Louisiana	18.6%	18.6%	18.6%	Tennessee	27.5%	27.5%	27.5%
Maine	42.2%	42.2%	42.2%	Texas	14.8%	14.7%	14.8%
Maryland	33.8%	33.8%	33.8%	Utah	14.4%	14.4%	14.4%
Massachusetts	16.4%	16.4%	16.4%	Vermont	25.7%	25.7%	25.7%
Michigan	23.8%	23.8%	23.8%	Virginia	34.1%	34.0%	34.1%
Minnesota	17.0%	17.0%	17.0%	Washington	7.6%	7.6%	7.7%
Mississippi	31.9%	31.8%	31.9%	West Virginia	47.5%	47.4%	47.5%
Missouri	21.5%	21.5%	21.5%	Wisconsin	20.9%	20.9%	20.9%
Montana	13.7%	13.6%	13.7%	Wyoming	19.1%	19.1%	19.1%

Table A12: Damages offset per additional dollar of government funds associated with marginal subsidy increases around the current system of subsidies. The first column ("Unit") gives the damages offset per dollar associated with marginal increases in unit-based subsidies, s_j^{Unit}, the second column ("Cost") gives the damages offset per dollar associated with marginal increases in cost-based subsidies, s_j^{Cost}, and the third column ("kWh") gives the damages offset per dollar associated with marginal increases in production-based subsidies, s_j^{kWh}.

	(1) No Demographics	(2) Add % College	(3) Add % Democrat	(3) Add % Homeowner
I. Δ Average Subsidy ($Thousands)				
Midwest	5.3	5.4	5.3	5.3
Northeast	0.4	0.5	0.4	0.4
South	4.7	4.7	4.7	4.6
West	0.3	0.4	0.4	0.4
II. Δ Installations per 1000HHs				
Midwest	2.9	2.9	2.8	2.8
Northeast	-10.1	-10.3	-9.6	-9.6
South	4.7	4.8	4.3	4.3
West	0.8	0.8	0.9	0.9
III. Δ Annual Damages Offset ($Millions)				
CO2e	6.8	7.2	6.1	6.1
NOx	1.4	1.5	1.3	1.3
PM2.5	0.4	0.5	0.3	0.3
SO2	1.3	1.2	1.3	1.3
Total	9.8	10.4	9.0	9.0

Table A13: Counterfactual results under alternative model specifications. Each entry shows the change of moving from the current system of subsidies to the optimal system of subsidies given the specification in question. The first panel shows the change in the average present discounted value of subsidies for a 15-panel installation for each census region. The second panel gives the change in the simulated number of solar installations per 1000 households in the model for each Census region. The final panel gives the change in total damages offset by rooftop solar. See text for details on each model specification.

plus a factor proportional to the square of flow on the line: $L_{Rt} = \alpha_{1R} + \alpha_{2R}\left(\text{Load}_{Rt} - E^{\text{Solar}}_{Rt}\right)^2$. Note that the parameters α_{1R} and α_{2R} are both allowed to vary by region to reflect differences in grid characteristics across regions. We then adjust excess load by those losses, $\text{ELoad}^{LL}_{Rt} = \text{Load}_{Rt} - E^{\text{NonD}}_{Rt} - E^{\text{Solar}}_{Rt} + L_{Rt}$. Losses enter positively since power plants must produce not only the amount of electricity demanded by households but also must make up for the losses incurred in transporting electricity to the household. Adding line losses changes the marginal damages offset by residential solar to

$$\left|\frac{\partial D\left(\text{ELoad}\right)}{\partial N_i}\right| = \sum_{t=0}^{T}\frac{1}{\left(1+r\right)^t}A_{it}\left(1 + 2\alpha_{2R}\left(\text{Load}_{Rt} + E^{\text{Solar}}_{Rt}\right)\right)\left|\frac{\partial D_t\left(\cdot\right)}{\partial \text{ELoad}_{Rt}}\right|.$$

The installation of solar panels now has two benefits. As we have in our primary model, solar panels reduce the electricity demand fulfilled by power plants, generating benefits equal to the electricity produced by a panel, A_{it}, times the change in damages, $\left|\frac{\partial D_t(\cdot)}{\partial \text{ELoad}_{Rt}}\right|$. Now, there is an additional benefit from offsetting line losses, captured by the term $2\alpha_{2R}\left(\text{Load}_{Rt} - E^{\text{Solar}}_{Rt}\right)$, which is the marginal change in losses. Including line losses increase the average damages offset by marginal installers, $\overrightarrow{\Delta D}^{\theta,\text{ext}}_{j}$, and by intensive margin installers, $\overrightarrow{\Delta D}^{\theta,\text{int}}_{j}$, when calculating optimal subsidies.

Borenstein and Bushnell (2022) estimate line losses as a proportion of total production for over 1,600 utilities in the United States. We take the weighted average of these estimates to create values for each region, weighting by the total electricity production of each utility. Let γ_R be line losses as a proportion of total production in region R. We then follow their assumption that 25% of line losses are independent of flow on the line, which allows us to back out $\alpha_1 = 0.25\gamma_R\sum_t\left(\text{Load}_{Rt} - E^{\text{Solar}}_{Rt}\right)$ and $\alpha_2 = 0.75\gamma_R\frac{\sum_t\left(\text{Load}_{Rt} - E^{\text{Solar}}_{Rt}\right)}{\sum_t\left(\text{Load}_{Rt} - E^{\text{Solar}}_{Rt}\right)^2}$.

Results The results are summarized in Table A14. The first column gives subsidies, installations, and damages offset given the current system of subsidies when we account for line losses. The annual damages offset are slightly larger than the baseline model in which we do not account for line losses.

The following summarize the results under (2) state-specific welfare-maximizing subsidies, (3) state-specific damage-minimizing subsidies, (3) nationally uniform welfare-maximizing subsidies, (4) nationally uniform damage-minimizing subsidies. In all four counterfactuals, the subsidies and installations are similar to those in the baseline model when we do not account for line losses. However, the environmental gains are slightly larger than in the baseline model.

	(1)	(2)	(3)	(4)	(5)
		State-Specific Subsidies		Tract-Specific Subsidies	
	Baseline	Welfare Max	Damage Min	Welfare Max	Damage Min
I. Average Subsidy ($Thousands)					
Midwest	11.3	16.7	17.1	16.7	17.0
Northeast	18.2	18.9	21.8	18.8	21.6
South	10.4	14.9	12.6	15.0	12.5
West	12.0	12.2	5.7	12.2	5.6
II. Installations per 1000HHs					
Midwest	4.1	6.9	8.6	7.0	8.9
Northeast	29.8	20.7	28.7	20.6	28.1
South	6.8	11.0	9.4	11.0	9.4
West	13.2	13.8	6.8	13.7	6.7
National	11.4	12.4	11.3	12.4	11.3
III. Annual Damages Offset ($Millions)					
CO2e	76.7	83.3	79.6	83.3	79.4
NOx	19.9	21.3	20.1	21.3	20.1
PM2.5	18.4	18.8	19.9	18.8	19.9
SO2	40.7	42.6	53.4	43.1	54.4
Total	155.7	165.9	173.0	166.4	173.9

Table A14: Optimal subsidies when accounting for line losses. The first panel shows the average present discounted value of subsidies received for a 15-panel installation for each census region. The second panel gives the simulated number of solar installations per 1000 households in the model for each Census region. Households are defined as rooftops which are suitable for solar panel installations as defined by GPS data. The final panel gives the total damages offset by rooftop solar.

D.3 Improved Storage of Nondispatchable Electricity

Because of intermittent nature of many renewable energy sources, times when renewable energy generation is high may not correspond with times when electricity demand is high. Improvements in energy storage technology would allow electricity generated by nondispatchable energy sources to be stored for times when it is most needed. What would be the environmental benefits of these improvements in energy storage technology? And how would the introduction of improved electricity storage technology change the optimal system of solar subsidies?

As a simple way to try to answer these questions, we consider a setting in which electricity produced by nondispatchable sources (including household solar) can be imperfectly reallocated over time. Specifically, given the total amount of electricity produced by nondispatchable sources in a year, we assume a proportion ω of this electricity is reallocated over time such that the profile of *usage* of this reallocated electricity is proportional to electricity demand.[78] Formally, we write excess demand as

$$\text{ELoad}_{Rt}^{\text{storage}} = \left(1 - \underbrace{\omega\left(A_R^{\text{NonD}} + A_R^{\text{Solar}}\right)}_{\text{Reallocated Electricity}}\right)\text{Load}_{Rt} - \underbrace{(1-\omega)\left(E_{Rt}^{\text{NonD}} + E_{Rt}^{\text{Solar}}\right)}_{\text{Non-Reallocated Electricity}}$$

where $A_R^{\text{NonD}} = \frac{\sum_t E_{Rt}^{\text{NonD}}}{\sum_t \text{Load}_{Rt}}$ and $A_R^{\text{Solar}} = \frac{\sum_t E_{Rt}^{\text{Solar}}}{\sum_t \text{Load}_{Rt}}$ are region-specific constants which ensure that total amount of nondispatchable energy utilized is equal to total nondispatchable energy generated.[79]

Results We calculate the environmental benefits of this improved storage technology and the optimal subsidies given the new storage technology for three values of ω in Table A15. Column (2), for example, shows the effects of this alternative storage technology with $\omega = .25$, holding the system of solar subsidies at their current levels. As subsidies do not change, the distribution of installations is the same as in the case without storage technology. Panel III shows that the improved storage technology leads to a decrease in environmental damages valued at over \$60 million annually. Column (3) recalculates the optimal subsidies given that the new storage technology is in place. We find that the optimal subsidies are very similar to the baseline case and that implementing the optimal subsidies leads to similar reductions

[78] This is highly stylized model of electricity storage. More generally, optimal storage and withdrawal of electricity will depend on the distribution of the cost of electricity production by other sources over time and space. See Holland, Mansur, and Yates (2022) for a richer model of electricity storage. It would be straightforward to only be reallocated within the same day it is generated.

[79] Similar to Holland, Mansur, and Yates (2022), we assume that there are no electricity losses associated with electricity storage, e.g. from charging batteries or decay of electricity over time.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		$\omega = .25$		$\omega = .5$		$\omega = .75$	
	Baseline	Improved Storage	+Optimal Subsidies	Improved Storage	+Optimal Subsidies	Improved Storage	+Optimal Subsidies
I. Average Subsidy ($Thousands)							
Midwest	4.1	4.1	6.8	4.1	6.8	4.1	6.7
Northeast	29.8	29.8	20.0	29.8	19.8	29.8	19.7
South	6.8	6.8	11.0	6.8	11.0	6.8	10.9
West	13.2	13.2	14.3	13.2	14.6	13.2	14.7
II. Installations per 1000HHs							
Midwest	11.3	11.3	16.6	11.3	16.6	11.3	16.6
Northeast	18.2	18.5	18.6	18.5	18.5	18.5	18.4
South	10.4	10.7	15.0	10.7	14.9	10.7	14.9
West	12.0	12.8	12.5	12.8	12.6	12.8	12.7
III. Annual Damages Offset by Stoarge and Rooftop Solar ($Millions)							
CO2e	69.8	94.1	100.3	111.7	117.9	122.5	128.7
NOx	18.0	36.2	37.5	51.1	52.5	62.7	64.1
PM2.5	16.7	28.3	28.6	36.9	37.2	42.3	42.5
SO2	37.0	46.6	47.6	55.2	56.0	62.8	63.4
Total	141.5	205.2	214.0	254.9	263.5	290.3	298.7

Table A15: Optimal subsidies with improved storage technology. The first panel shows the average present discounted value of subsidies received for a 15-panel installation for each census region. The second panel gives the simulated number of solar installations per 1000 households in the model for each Census region. The final panel gives the total damages offset by rooftop solar and by the increased storage technology of renewable energy.

in environmental damages as we find without the improved storage technology.

The remaining columns repeat this exercise for $\omega = .5$ and $\omega = .75$. In both scenarios, we find large environmental benefits to the new technology. However, the optimal subsidies and the environmental benefits associated with implementing those subsidies are very similar to those in the baseline case.

D.4 Cleaner Commercial Production

We present our results when we allow changes in commercial energy production in Table A16. The first column gives the results under the current technology, as in our baseline results. The first panel gives the change in average subsidy, measured in thousands of dollars when moving from the current subsidies to the welfare-maximizing subsidies. The second panel gives the change in installations per 1000 households. The final panel gives the percentage change in environmental benefits when moving from the current to welfare-maximizing subsidies. As before, we can see that moving to the optimal subsidies given the baseline technology leads to an increase in environmental benefits of 6.4%.

The next three columns show the results when we recalculate welfare-maximizing subsidies given that the scale of utility-scale solar and wind expand based on three scenarios of

	(1)		(2)	(3)	(4)
	Current	Increased Renewables			Clean
	Tech	High Cost	Mid Cost	Low Cost	Coal
I. Δ Average Subsidy ($Thousands)					
Midwest	5.3	6.0	6.1	6.3	3.9
Northeast	0.4	0.8	0.8	0.8	-0.2
South	4.7	5.1	5.1	5.1	4.7
West	0.4	-0.5	-0.5	-0.5	1.1
II. Δ Installations per 1000HHs					
Midwest	2.8	3.3	3.4	3.5	1.7
Northeast	-9.6	-8.7	-8.7	-8.8	-10.4
South	4.3	4.8	4.8	4.8	4.2
West	0.9	-1.1	-1.2	-1.2	2.4
					1.1
III. %Δ Environmental Benefits					
Total	6.4	10.0	10.8	10.8	5.2

Table A16: Counterfactual results under alternative assumptions about commercial energy production. Each entry shows the change of moving from the current system of subsidies to the optimal system of subsidies given the specification in question. The first panel shows the change in the average present discounted value of subsidies for a 15-panel installation for each census region. The second panel gives the change in the simulated number of solar installations per 1000 households in the model for each Census region. The final panel the percent increase in aggregate environmental benefits. See text for details on each model specification.

projected renewable expansion by 2030 from the EIA (Nalley and LaRose, 2022). Specifically, we expand wind and solar based on their "reference case" projection, low-cost projection, and high cost-projection. The high-cost scenario is associated with the smallest increase in utility-scale solar and wind production, while the low-cost scenario is associated with the largest increases.[80] We refer to their reference case projection as the mid-cost projection. Across the three scenarios, we find that moving to the optimal subsidies leads to a 10.0-10.8% increase in aggregate environmental benefits.

In the final column, we recalculate results considering each coal plant to have "cleaned up" by adjusting marginal damages from coal plants so that the distribution of marginal damages from coal plants matches that of natural gas plants. Moving to the optimal subsidies leads to a 5.2% increase in aggregate environmental benefits in this case.

[80]Specifically, utility-scale solar increases by roughly 200%, 350%, and 500% in the three scenarios, while wind increases by 45%, 50%, and 55%.