Optimal Subsidies for Residential Solar*

Mark Colas
University of Oregon

Emmett Reynier
University of Oregon

September 16, 2024

Abstract

We study the optimal design of spatially differentiated subsidies for residential solar panels. We build a structural model of solar panel demand and electricity production across the US 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. National funding for subsidies under the current system exceeds the unconstrained optimum by over 70%. Our results suggest that there could be large welfare gains to redistributing funds towards other programs.

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 Manuel Amador, Severin Borenstein, Jon Davis, David Evans, Doireann Fitzgerald, Stephie Fried, Simon Fuchs, Joshua Graff Zivin, Kyle Herkenhoff, Illenin Kondo, Ashley Langer, Keaton Miller, Simon Mongey, John M. Morehouse, Kathleen Mullen, Nick Muller, Ed Rubin, Jim Schmitz, Chris Timmins, Woan Foong Wong, Eric Zou, and seminar participants at the University of Oregon and the Minneapolis Fed. This work benefited from access to the University of Oregon high performance computing cluster, Talapas. This paper was previously titled "Optimal Solar Subsidies."

1 Introduction

Governments around the world are more likely to subsidize technologies that lead to environmental benefits than they are to tax environmental damages (Borenstein, 2012). In the case of solar panels, these environmental benefits vary geographically. All else equal, the environmental benefits of solar panels are likely to be the largest in sunny areas and areas where high-polluting power plants would otherwise produce electricity. This spatial variation in environmental benefits suggests a rationale for spatially differentiated subsidies.

This paper uses a structural model of electricity production and household demand for rooftop solar panels to study how subsidies should optimally vary across space. 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.[1] 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. 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 nationally-optimal subsidies and quantify the benefits of switching from the current system of subsidies to the optimal subsidies.[2] Doing so requires understanding how solar panel installation rates and the damages associated with electricity production would change in response to alternative subsidy schemes. Therefore, our approach

[1] See Borenstein (2017) for a detailed discussion of the private benefits of solar installation in the case of California. See Borenstein and Bushnell (2022) for a discussion of how these factors shape the spatial distribution of solar panels.

[2] We focus on the optimal choice of subsidies for rooftop solar panels and do not allow for other types of government intervention, such as pricing the externality via a carbon tax. See Eichner and Runkel (2014) for an argument for why countries may choose to subsidize green energy production even when they have access to carbon taxes.

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 novel data sources to estimate the household component of the model via indirect inference, thereby providing the first estimated model of solar panel demand across the US. To facilitate identification, we utilize a *border discontinuity* approach, which exploits variation in subsidies on either side of state borders. 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 to solar rebates.

To model power plants, we develop a novel policy function approach that maps electricity demand and renewable production across the country to plant-level electricity production and emissions. Our approach allows for endogenous changes in power plants' production profiles in response to electricity demand and renewable production over the day and year. We estimate these policy functions using Open Grid Emissions (OGE) data, which provide hourly production and emissions data covering nearly every power plant in the United States. We show that the estimated model matches the data's temporal and spatial distribution of electricity generation. We translate these emissions into environmental damages using AP3, a state-of-the-art integrated air pollution model.

Our estimated model of solar panel demand and electricity production provides a framework to calculate the spatial distribution of installations, environmental benefits of solar panels, and government cost of subsidies under counterfactual subsidy schemes. We first use this framework to solve for the nationally-optimal cost-neutral subsidy reforms and quantify the spatial 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 100% relative to the nationally-optimal subsidy system, leading to 120% greater installations than optimal. 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 panels are under-subsidized by nearly 70%, leading to installations that are 60% lower than optimal. More generally, panels are under-allocated by roughly 30% in the Midwest and South and over-allocated by nearly 80% in the Northeast.

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 11.5% increase in aggregate environmental benefits—environmental damages decrease by approximately the same amount as a 11.5% 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 a 12.6% increase in aggregate environmental benefits.

Next, we calculate nationally-optimal subsidies when the government does not face an externally set budget constraint. The optimal unconstrained subsidies are significantly lower than current levels in the Northeast and West, but slightly higher than current subsidies in the Midwest and South. Nationally, total installations under the optimal subsidies are roughly 20% less than the current amount, leading to a decrease of 43 million dollars in annual environmental benefits relative to the current level. However, the accompanying 162 million dollar annual decrease in government costs thoroughly outweighs the decrease in environmental benefits. Put another way; the optimal unconstrained subsidies achieve over 80% of current environmental benefits at less than 60% the current cost. Our results suggest rooftop solar subsidies not only deviate from the optimum in how they vary across space but are also excessively generous in general.

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 over 3.5 times greater than the damages offset per dollar of subsidy increases in Washington. These results highlight that changes around the current system of subsidies could lead to decreases in both environmental damages and fiscal costs.

The remainder of the paper consists of various extensions and robustness checks. We analyze the sensitivity of our results to 1) alternative values of the marginal cost of public funds, 2) alternative specifications of household preferences and discounting, 3) accounting for line losses in transmitting electricity from plants to homes, 4) accounting for transmission constraints, 5) the introduction of improved electricity storage technology, and 6) changes in utility-scale renewable electricity production. We find that the optimal system of subsidies remains qualitatively the same across these specifications. Quantitatively, our results suggest that optimal unconstrained subsidies will be even lower in the future as utility-scale renewable electricity production continues to expand.

It is important to caveat that residential solar subsidies may be associated with additional benefits that we do not model in this paper. Solar subsidies may increase innovation and technological growth via increased R&D or learning-by-doing (Benthem, Gillingham, and Sweeney, 2008; Goulder and Mathai, 2010; Bollinger and Gillingham, 2019; Gerarden, 2023).[3] Further, numerous studies have found that solar panel suppliers hold considerable market power (Gillingham et al., 2016; Pless et al., 2017; Pless and Van Benthem, 2019). Subsidies may help to alleviate the distortions caused by this market power (Dorsey, 2024). Solar installations may encourage additional solar installation via peer effects (Bollinger and Gillingham, 2012; Bollinger et al., 2022) and may reduce reliance on the transmission grid, thereby reducing grid congestion and the need for costly investments in the transmission grid. Although not accounted for, these benefits would need to be substantial in order to justify current subsidy spending levels.

Further, the goal of this paper is to quantify *nationally-optimal* subsidies. Specifically, we focus on a social planner who maximizes national welfare, accounting for aggregate environmental damages and national fiscal costs. In reality, many solar subsidies are set by state governments who face state-level budget constraints and may be motivated to create local jobs or stimulate the local economy via "green multipliers" (Popp et al., 2020; Hasna, 2021; Batini et al., 2022). In our analysis, we abstract away from these state-level incentives and constraints. Our results quantify the theoretically optimal distribution of funds across space and the geographic misallocation associated with the suboptimal current system of subsidies.

Our paper is most closely 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 various subsidy schemes. Of these, our paper is closest to Feger, Pavanini, and Radulescu (2022), which studies optimal installation subsidies and energy tariffs in a model with household energy consumption and solar panel demand.[4] While all these papers focus on solar panel installations, we provide a framework that can additionally quantify the environmental benefits of solar panel installations, arguably the main reason these subsidies

[3] Learning by firms only constitutes a positive externality if it is non-appropriable and therefore not internalized by the firm. Bollinger and Gillingham (2019) find that non-appropriable learning-by-doing only leads to small learning spillovers for solar panel installers in California.

[4] Feger, Pavanini, and Radulescu (2022) quantify the cost-minimizing and welfare-maximizing subsidy and tariff schemes subject to a network financing constraint and a solar energy target. They also allow the government to have a preference for redistribution, something we refrain from doing in our paper.

exist. As such, we are the first paper in this literature to quantify the trade-offs between the environmental benefits and fiscal costs of residential solar subsidies. We additionally contribute by quantifying the spatial misallocation due to current subsidy schemes through our analysis of how these subsidies should optimally vary across space. As discussed in the following paragraph, the reduced-form literature has emphasized the importance of spatial differences in the environmental benefits of solar panels. However, no quantitative research has incorporated these spatial differences in a study of optimal subsidy design.

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), Siler-Evans et al. (2013), Graff Zivin, Kotchen, and Mansur (2014), Holland et al. (2016), Millstein et al. (2017), Callaway, Fowlie, and McCormick (2018), Holland et al. (2020), Brown and O'Sullivan (2020)). In particular, this paper is similar to Borenstein and Bushnell (2022), who relate estimates of the marginal social cost of electricity production to the spatial distribution of solar panels in the US, and to Sexton et al. (2021), and Lamp and Samano (2023), who study the marginal benefits of solar panel installations and calculate the benefits of reallocating panels across space.[5] While these papers establish that the current spatial distribution of solar panel installations does not maximize environmental benefits, they do not model demand for solar panels and, therefore, do not quantify how installations respond to various subsidy schemes. The goal of our paper is to quantify the extent to which government policy causes this misallocation and solve for the system of subsidies that remedies this misallocation. 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 and fiscal costs. Specifically, we provide the first estimated model of solar panel demand across space in the US. We also develop a novel, tractable approach to modeling power plant production and the associated emissions over space and time. This approach involves directly modeling how individual plants' electricity production and emissions endogenously respond to changes in solar and other renewable production.

Finally, this paper is related to several empirical papers which estimate the elasticity of solar panel installations with respect to subsidies. We discuss these papers in Section 5.1.3. We use the estimates from these papers to evaluate the performance of our estimated model.

[5] In the sustainability literature, Tibebu et al. (2021) calculate the subsidies which maximize environmental benefits less government cost at the national and state level. Their analysis does not account for household utility and therefore omits a key component of the social benefit of subsidies. They also do not model the household decision to install solar panels but instead model solar installation rates as following a normal distribution in the net present value of installation.

2 Model

We combine a model of household solar panel demand with an 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, central generation power plants 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.[6] 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. 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 household i's energy consumption in period t. We assume the household pays a constant price of p_j for all electricity purchased.[7] 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.

[6] 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)).

[7] 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.

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. The installation cost function $p_j^{\text{Ins}}(\cdot)$ can vary nonlinearly in N_i and is allowed to vary by state j.[8] 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, as is the case for households in states with net metering. We discuss how we model households without net metering in Online Appendix B.1.[9] 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. 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^{Panel}, such as subsidies that pay per kilowatt of solar capacity installed.

We can thus write the household's budget constraint as

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

where c_i is consumption of the numeraire good.[10]

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.

Note that the choice of electricity consumption does not depend on the household's choice to install panels. Thus, we can think of household optimization as a two-step process. First,

[8] 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.

[9] 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.

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

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

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

denotes household i's net monetary benefit of installing solar panels.[11] 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.[12] As we show in Section 2.3, the planner chooses the nationally-optimal set of subsidies accounting for the fact that different households are marginal with respect to each type of subsidy.

Before proceeding, it is important to highlight that several features of the solar panel market are not included in our model. First, our model does not include a role for peer effects, which Bollinger and Gillingham (2012) and Bollinger et al. (2022) find can influence households' decisions to install solar panels.[13] Second, De Groote and Verboven (2019) find that households heavily discount future benefits associated with solar installations. We instead assume households discount future subsidy payments according to the market discount rate. We consider a version of our model in which households more heavily discount future subsidy payments in Section 7.2. Third, we assume a partial equilibrium setting where installation costs are given exogenously and do not change in response to changes in subsidies. This partial equilibrium setting implies that increases in subsidies decrease the effective costs paid by consumers dollar for dollar. If the inverse supply curve for solar installations in our model were upward sloping, then increasing subsidies would also increase the equilibrium

[11] 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.

[12] Note that electricity prices and production subsidies have the same effect on the monetary benefits from installation for households in states with net metering.

[13] Our estimation strategy uses tract-level data to estimate the effects of solar subsidies on installations. Tract-level responses to subsidies include both the direct effect of subsidies on installations as well as the effect of within-tract peers on total installations. Therefore, the empirical elasticities we use in our analysis incorporate the effects of peer effects on installations. However, we do not model these peer effects directly.

price charged by installers, thereby changing the price faced by the consumer.[14]

Finally, our paper assumes prices and subsidies are fixed over time and, therefore, abstracts away from the substantial changes in residential solar subsidies that have occurred over the past 20 years in the United States (Barbose and Darghouth, 2019). We view our model of solar installations as capturing the *long-run* installation decisions of households who face a given set of subsidies. As such, our model is not appropriate for analyzing the short-run effects of policy changes nor for studying the transition path between two policy regimes. We discuss the implications of assuming constant prices for estimation and quantitative results in Section 7.8.

2.2 Electricity Production

2.2.1 Background

Before proceeding to the model, we give a brief overview of 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 central generation 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. 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 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.[15] This implies that a

[14] See Reguant (2019) for a model-based exploration of how various renewable energy policies differentially affect the prices faced by consumers.

[15] 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

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

2.2.2 Model: Electricity Production

Within the model, three sources supply electricity: 1) residential solar, 2) nondispatchable plants, and 3) dispatchable plants.[16] Nondispatchable units are assumed to operate at full capacity conditional on environmental conditions (e.g., sun and wind) and conditional on total demand exceeding the amount produced by these nondispatchable generators.[17] Therefore, as long as demand exceeds the amount produced by nondispatchable sources, the production by these power plants is independent of demand and production by other plants. Alternatively, the production by dispatchable units depends on excess demand remaining after production by residential solar and nondispatchable plants.

Residential Solar and Nondispatchable Plants Let R index NERC regions.[18] Total residential solar production in region R in a given hour t is the sum of energy produced by residential solar panels, $E_{Rt}^{\text{Solar}} = \int_{i \in I_R} m_i^\star N_i^\star A_{it} di$, where 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.

higher variable costs (such as gas-fired plants) begin operating only when electricity demand is sufficiently high.

[16] 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 7.6, 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.

[17] In this case, we assume production of nondispatchable plants is curtailed such that supply does not exceed demand.

[18] 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 Online Appendix A.7.

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 the total electricity demand in region R in time t, and let $\text{ELoad}_{Rt} = \text{Load}_{Rt} - E_{Rt}^{\text{NonD}} - E_{Rt}^{\text{Solar}}$ 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_{kt}^{Disp} denote production by dispatchable plant k in time t, we specify

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

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}.[19] We allow the function $f_k(\cdot)$ to differ across plants to reflect heterogeneity in the order in which plants are dispatched. We also allow $f_k(\cdot)$ to depend not only on excess load in a power plant's own region 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.

Intuitively, y_{kt}^{Disp} captures how production by an individual power plant k in a given hour t responds to fluctuations in electricity demand and nondispatchable production across the grid. For example, as the sun goes down and solar production decreases, excess load will increase across the country, particularly in regions heavily reliant on solar energy. y_{kt}^{Disp} tells us how individual power plants are dispatched to match these increases in excess load.

2.2.3 Damages

Let $d_k\left(y_{kt}^{\text{Disp}}\right)$ 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. We assume these damage functions are given exogenously and, therefore, rule out the possibility that power plants endogenously switch fuel types in response to policy changes. Let $D_t\left(\text{ELoad}_t\right) = \sum_k d_{kt}\left(y_{kt}^{\text{Disp}}\right)$ denote total damages from all power

[19] 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 or if plants face significant ramping constraints. We can accommodate this extension by allowing the function $f_k(\cdot)$ to depend on lagged values of excess demand, or on lagged production levels of the individual plant.

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

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

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 Nationally-Optimal Subsidies

The government chooses subsidies to maximize the sum of total utility minus total environmental damages subject to an externally set budget constraint.[20] We consider a government who does not face a budget constraint in Section 6.4. To ease up on notation, let $s_{ij} = s_j^{\text{Panel}} N_i^\star + s_j^{\text{kWh}} A_i N_i^\star + s_j^{\text{Cost}} p_j^{\text{Ins}}\left(N_i^\star\right)$ 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 $\theta \in \{\text{kWh}, \text{Panel}, \text{Cost}\}$, 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 $M_j = \int_{i \in I_j} m_i di$ denote the total number of households who install solar panels in state j, where I_j is the set of households in state j.

The government maximizes the sum of utility less environmental damages,

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

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

[20] 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.

set constraint

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

where G is the maximum amount the government can spend on subsidies. We can reformulate the government's objective function as the Lagrangian

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

where λ is equal to the marginal cost 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 nationally-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 Online Appendix B.3 and provide definitions for each individual object in the upcoming text. The first term ("Extensive Margin") captures the trade-off between environmental benefits and fiscal costs associated with households who are *additional* with respect to a small subsidy increase: 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

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

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

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

where ELoad^{SB} is the excess load under the optimal system of subsidies. These additional 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

$$\overrightarrow{s}_j^{\theta,\text{ext}} = \frac{\int_{i \in I_j} s_{ij} \overrightarrow{m_i}^{\theta} di}{\int_{i \in I_j} \overrightarrow{m_i}^{\theta} di}.$$

The second term of equation (8) ("Intensive Margin") captures the environmental-fiscal trade-offs 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

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

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 additional panels.[21] Taken together, these first two terms show that the government will optimally increase subsidies which induce a greater number of additional installations and additional panels from households associated with significant environmental benefits and for whom fiscal costs are low.

The final term ("Mechanical Effect") captures the effects of increasing subsidies for the *non-additional* households: 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.[22] Each dollar transferred to these non-additional households increases welfare by $(1 - \lambda)$, which reflects the increase in household utility less the decrease in government funds.[23] In summary, equation (8) measures the effects of subsidy changes on welfare, accounting for environmental benefits, fiscal cost, and household utility.

However, solving for nationally-optimal subsidies requires more structure on the problem.

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

[22] Formally this is $\overline{\frac{\partial s_{ij}}{\partial s_j^{\theta}}} = \frac{\int_i m_i^{\star} \frac{\partial s_{ij}}{\partial s_j^{\theta}} di}{\int_i m_i^{\star} di}$.

[23] Note that the utility of additional households does not show up in equation (8) since there is no first-order welfare effect for households who are additional with respect to a marginal subsidy increase (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.

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. 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 non-additional 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 nationally-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 and fiscal costs.

3 Quantitative Model

3.1 Household preferences

Let ℓ denote the census tract in which a household lives. We assume that the nonpecuniary utility of installation of installing N_i panels, $\gamma_i(N_i)$, 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

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

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. 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 percent of voters who voted Democrat in the 2016 presidential election.

Recall that the number of panels installed cannot exceed the space the household has available for panels, denoted by $\bar{N}_i$. The optimal number of panels conditional on installation is therefore given by

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

where $\frac{\partial \mu_{ij}}{\partial N_i}$ is the derivative of monetary benefit of installations (μ_{ij}) with respect to installation size (N_i). 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.[24] 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

$$\frac{\partial \log\left(\pi_i\right)}{\partial \mu_{ij}\left(N_i^{\star}\right)} = \frac{1}{\sigma}\left(1-\pi_i\right). \tag{11}$$

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

Production by dispatchable power plant k in time t is given by equation (4). We assume the latent function $f_k\left(\text{ELoad}_t, \varepsilon_{kt}\right)$ is a quadratic 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

$$f_k\left(\text{ELoad}_t, \varepsilon_{kt}\right) = \psi_k^0 + \sum_{R\in\mathbf{R}_k}\left(\psi_{Rk}^1 \text{ELoad}_{Rt} + \psi_{Rk}^2 \text{ELoad}_{Rt}^2\right) + \varepsilon_{kt}, \tag{12}$$

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 transmit-

[24] As we explain below, we parameterize $p_j^{\text{Ins}}\left(N_i\right)$ 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 $N_i > 0$.

ted across regions within interconnections but is rarely transmitted across interconnections.[25] 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.

Further, while the latent function $f_k(\cdot)$ is assumed to be constant across time, we show in Section 5.2 that our model can replicate differences in dispatchable production over the day and year in response to fluctuations in nondispatchable production and electricity demand. In particular, our model can generate the ramping pattern of dispatchable generators through the afternoon as solar generation decreases and electricity demand increases. An important caveat to our approach is that these plant-specific policy functions are not invariant to changes in factors which may change the order in which plants are dispatched, such as changes in fuel costs or the introduction of a carbon tax. However, we do not expect first-order effects on these factors in the counterfactual subsidy schemes we investigate.

It is worth discussing how the specification of power plant production we develop here differs from the specifications used in Holland et al. (2016) and Sexton et al. (2021). Those papers estimate marginal emissions rates for individual power plants in which time-specific reduced-form coefficients capture all differences in emissions rates across time. These specifications, therefore, do not model how production levels endogenously respond to fluctuations in renewable production. As such, the estimated emissions rates for each power plant are constant conditional on time, and independent of the amount of solar electricity produced.

Since we aim to estimate marginal emissions both under current conditions and under significant changes to the distribution of residential solar panels, we require a different approach to modeling power plants. In our model, production varies flexibly in excess load and therefore is endogenous to both electricity demand and production from solar and other re-

[25] 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.

newable sources. 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 $g \in \mathcal{G}$ index pollutants, where we assume the set $\mathcal{G}$ consists of the pollutants NOx, PM2.5, SO2, and CO2 equivalent (CO2e).[26] As shown by Holland et al. (2022), power plants' marginal emissions rates tend to decline as utilization increases, where utilization is a plant's production level relative to its capacity. Therefore, to allow for emissions rates which vary across production levels, 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

$$\text{Emis}_{gkt}(y_{kt}) = \begin{cases} \kappa_{gk}^1 y_{kt} + e_{gkt} & \text{if } (y_{kt} - y_k^{50}) < 0 \\ \kappa_{gk}^1 y_{kt} + \kappa_{gk}^2 (y_{kt} - y_k^{50}) + e_{gkt} & \text{if } (y_{kt} - y_k^{50}) \geq 0 \end{cases}. \tag{13}$$

Power plant k's damages in time t are then given by $d_{kt}(y_{kt}) = \sum_{g \in \mathcal{G}} \delta_{gk} \text{Emis}_{gkt}(y_{kt})$, where δ_{gk} gives the marginal damages associated with emissions of g by power plant k, accounting for power plant k's location and stack height.

[26] 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.

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 Online 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 from the year 2016.[27] 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.[28]

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 county-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 0.5% each year.[29] 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

[27]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.

[28]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.

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

of electricity production with panel specifications and climate as the model inputs (Blair, Dobos, and Gilman, 2013). For each county, we calculate the fraction of yearly solar production produced at any given hour over the year. See Online 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 to calculate state and federal subsidies in 2017. For the price of electricity, we use the average retail price of electricity as reported by the EIA.[30] We use a value of $r = 2\%$ for the real interest rate.

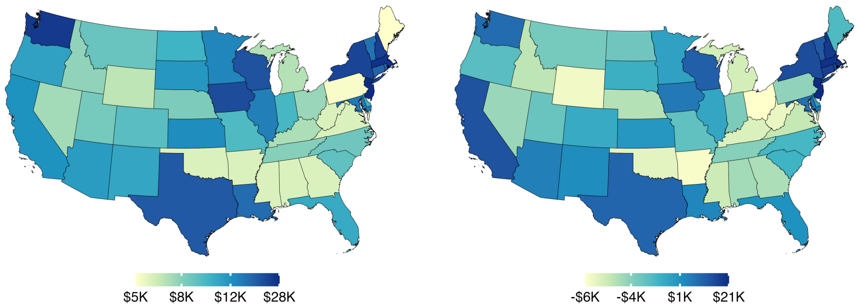
We estimate installation prices using data from Tracking the Sun, a project collecting data on solar panel installations by the Lawrence Berkeley National Lab. As Tracking the Sun only covers 25 states, we assume that all states within a given Census region share the same installation pricing function. 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 Online 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 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.[31] 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 nearly 10,000 power plants. 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

[30] These data can be downloaded at `https://www.eia.gov/electricity/state/`. See Ito (2014) for evidence that consumers respond to average, rather than marginal, electricity prices.

[31] We use data from 2019 as it is the first year available from OGE and thus closest to the Deepsolar data while also reflecting the modern electricity grid. The OGE methodology relies on the EIA form 930, which is only available starting in mid-2018.



(a) Expected subsidy

(b) Expected monetary benefit

Figure 1: Expected subsidies and monetary benefit for a 15-panel system in each state. Colors are scaled by the percentile of their respective value. Subsides and monetary benefits are measured in 2014 dollars. See text for details.

data. The plants excluded from CAMD data account for nearly 30% of NOx emissions, 8% of SO2 emissions, and 7% 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 Online 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.[32] AP3 and its predecessors, APEEP and AP2, have been employed extensively in the environmental economics literature.[33] In addition to AP3, we use the social cost of carbon to quantify damages from greenhouse gas emissions.

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. There is considerable variation across states in the generosity of these subsidies. New Jersey delivers nearly 29 thousand dollars in subsi-

[32] 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.

[33] See, e.g., Muller, Mendelsohn, and Nordhaus (2011), Holland et al. (2016), Shapiro and Walker (2020), Holland et al. (2020), Sexton et al. (2021), Cicala et al. (2021), Holland et al. (2021).

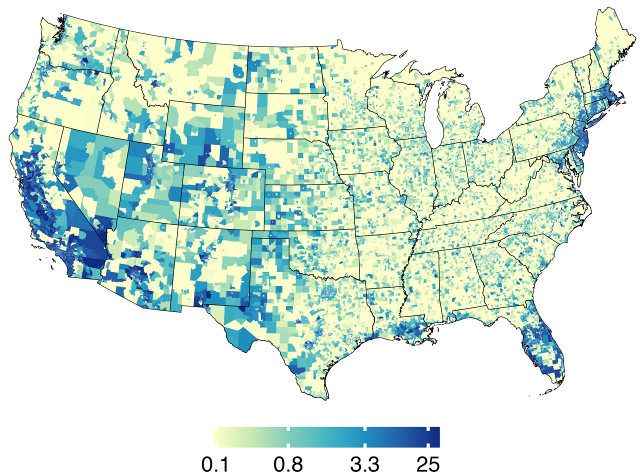
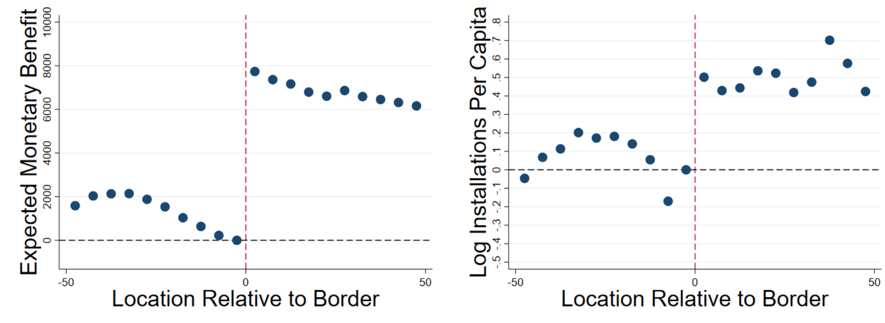


Figure 2: Installed solar systems per 1000 individuals.

dies, compared to seven states providing no additional funding, leading to under 6 thousand dollars in subsidies from the federal government. For comparison, the average cost of a solar installation in 2017 was roughly 24.5 thousand dollars. Therefore, the present discounted value of subsidies ranges from roughly one-quarter to more than the entire cost of an average installation. Figure A4 in Online 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 panel-based, subsidies.

In addition to subsidies, the monetary incentives to install panels vary geographically because of spatial differences in prices and sunlight. 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 2 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. Meanwhile, households in the Midwest, where there are relatively low subsidies, less sunlight, and low electricity prices, install few solar panels.

In Appendix C.3, we investigate how the monetary benefits of solar installations corre-



(a) Expected Monetary Benefit

(b) Log Installations Per Capita

Figure 3: Border Discontinuities in Monetary Benefits and Installations. Each graph plots estimated location-bin fixed effects from a regression of the variable in question on border and location-bin fixed effects. Monetary benefits are measured in 2014 dollars. Positive values on the X-axis represent households on the side of the border with more generous subsidies, and negative values on the X-axis indicate the side of the border with less generous subsidies.

late with the average size of solar installations. Across a number of specifications, we find that increased monetary benefits are associated with a statistically significant, but small in magnitude, increase in the average size of installations.

Border Discontinuities When we estimate our structural model, we use a border discontinuity approach, which compares installation rates on either side of state borders. Here, we present descriptive evidence on how the monetary benefits of installation and installation rates change as we cross the border from states with relatively less generous subsidies to states with more generous subsidies.

For this exercise, we limit our sample to tracts within 50 miles of state borders. Given the average solar irradiance in the border region, we calculate the net present value of subsidies a household would receive for a 15-panel installation. Then, we classify the side of the border for which this hypothetical subsidy is higher as the "generous" side. Finally, we divide households into 5-mile-wide bins in locations relative to the state border. To examine how a given variable changes as we approach and cross state borders, we follow Bayer, Ferreira, and McMillan (2007) and regress the variable in question on state-border fixed effects and location-bin fixed effects. We plot these estimated location-bin fixed effects, which give the conditional average of the variable in question in a given location bin relative to the bin nearest to the border on the less generous subsidy side (the omitted location-bin fixed effect).

Figure 3a shows how the expected monetary benefit varies by the location relative to the border. As before, we measure monetary benefit as the average net present value of installing a 15-panel system in a given tract. Positive values on the X-axis represent households on the side of the border with more generous subsidies, and negative values on the X-axis indicate the side with less generous subsidies. By construction, tracts on the negative (left) side of the border have lower monetary benefits, with a jump of roughly \$8,000 as we move from the less generous side of the border to the more generous side. Figure 3b shows the results for log installation rates. There is a sharp increase in installation rates across the border—moving from the less generous side of state borders to the more generous side is associated with roughly a 50% increase in installations per capita.

As argued by Black (1999), Bayer, Ferreira, and McMillan (2007) and Lee and Lemieux (2010), these discontinuities in installation rates are informative about the causal effect of subsidies on installations if other variables which may influence installation rates are continuous at state borders. One particular concern is that household preferences for solar panels may be discontinuous at state borders. This discontinuity would occur if, for example, households with stronger preferences for solar panels sorted onto the side of the state border with more generous solar subsidies. Though we cannot measure preferences for solar panels directly, in Online Appendix C.4, we look for suggestive evidence for this type of sorting by plotting 1) the percent of households with college degrees, 2) the percent of voters who voted Democrat in the 2016 presidential election, and 3) average household income as a function of distance to state border. We find no evidence of sorting around state borders among these characteristics.

Another concern is that differences in other state policies, such as state tax rates, could drive these discontinuities. To assess this concern, we run border discontinuity regressions of log installations per capita on subsidies for solar panels, both with and without controls for various state taxes. Table 1 shows the results. For each column, we limit our sample to tracts within 10 miles of state borders and regress tract-level log installations per capita on the net present value of subsidies for a 15-panel system, tract-level household demographics, and state-border fixed effects.[34] Column (1) gives regression results without any controls for state and local taxes. Column (2) adds the state income tax rate, measured as the average state tax rate evaluated at a household income of 60,000 dollars, roughly the average income in the data. Column (3) adds sales tax, measured as the average state and local sales tax in a given state, and Column (4) adds the average property tax rate in the state.[35] The estimates from these border discontinuity regressions are similar across all specifications, but the parameter estimate slightly increases as we add state tax controls. The estimate in Column (4) suggests a \$1,000 increase in subsidies is associated with roughly an 8% increase in installations. We will control for these state tax variables in the border discontinuity

[34]We include fixed effects for all state pairs which share a border. For example, California-Oregon, California-Nevada, and California-Arizona are all included as separate fixed effects.

[35]Sales tax data are taken from Walczak and Drenkard (2017). Property tax rates are from Sexton et al. (2021), who estimate property tax rates using data on real estate tax payments and property values from the American Community Survey.

	Dependent Variable: Log Installations per Capita			
	(1)	(2)	(3)	(4)
NPV Subsidies ($1000s)	0.0629***	0.0755***	0.0751***	0.0826***
	(0.00920)	(0.00438)	(0.00570)	(0.00761)
Income Tax	NO	YES	YES	YES
Sales Tax	NO	NO	YES	YES
Property Tax	NO	NO	NO	YES
Distance Bandwidth	10	10	10	10
Observations	6,052	6,052	6,052	6,052

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

Table 1: Regression of log installations per capita on the net present value of subsidies for a 15-panel installation within 10 miles of state borders. Subsidies are measured in thousands of 2014 dollars. State-clustered standard errors in parentheses. All regressions contain border fixed effects and tract-level demographic controls.

regressions we use in our structural estimation procedure.

Moreover, apart from the policies categorized as subsidies, there exist additional state measures designed to promote solar adoption. If the prevalence of these policies correlates with subsidy generosity, this may lead us to overestimate the responsiveness of installations to subsidies. In Appendix C.5, we rerun these border discontinuity regressions with additional controls for other state-level programs aimed at increasing solar installations that our subsidy measures do not account for. Our results are robust to the addition of these additional controls.

4.3 Estimation

Households We estimate the household installation component of the model via indirect inference. In essence, we first compute a set of "auxiliary models" that describe installation behavior in the data, and then simulate the structural model and calculate the auxiliary models with simulated data. We choose the six structural parameters $\sigma, \gamma_0, \gamma_{Coll}$, γ_{Pol}, γ_{1N} and γ_{2N} such that the auxiliary models computed from the model are as close as possible to those from the data.

Our first auxiliary model is a border discontinuity regression which measures how installation rates change as we cross state borders.[36] Formally, we run the following regression

[36] Similar border discontinuity approaches have been used frequently in the environmental literature (see, e.g., Ito (2014), Feger, Pavanini, and Radulescu (2022), Hughes and Podolefsky (2015), or Rubin and Auffhammer (2023)).

using only tracts within 10 miles of state borders:

$$\log M_\ell = \alpha \bar{\mu}_\ell + \beta Z_\ell + \theta_{border(\ell)} + \varepsilon_\ell,$$

where M_ℓ is the total number of solar installations in tract ℓ, $\bar{\mu}_\ell$ is the average monetary benefit of installing a 15-panel system in tract ℓ, β is a vector of regression coefficients, Z_ℓ is a vector of tract-level demographic and tax rate controls, and $\theta_{border(\ell)}$ is a state-border fixed effect.[37] The parameter α measures the relationship between installation rates and monetary benefits in narrow bandwidths around state borders, controlling for border fixed effects. We target the coefficient α as an auxiliary model parameter.[38] As argued above, α is informative of the causal effect of monetary benefits on installation, given that household characteristics potentially correlated with preferences for solar do not exhibit a discontinuity at state borders. Additionally, we also target 1) log installations per household in each census tract, and 2) the average number of panels per array in each census tract.

The six structural parameters of interest are well identified. As we can see from (11), the parameter σ dictates the extent to which installations increase with monetary benefits. This parameter is thus identified by the coefficient α from the border discontinuity regression. Variation in demographics across tracts then jointly identifies γ_0, γ_{Coll}, and γ_{Pol}. 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 (12) via maximum likelihood. We provide the likelihood function and additional details in Online 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. We estimate the damages given by equation (13) 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

[37] We again include fixed effects for all state pairs which share a border. The vector of controls Z_ℓ includes the tract-level college completion percentage and percent of voters who voted Democrat in the 2016 presidential election, the state income tax rate, the sales tax rate, and the average property tax rate.

[38] Our estimates of α are not sensitive to the size of the bandwidth around state borders, nor the inclusion of demographic controls. We obtain similar estimates of structural parameters when we instead include a regression of log installations on subsidy levels rather than total monetary benefits.

		Estimate	Standard Error
Dispersion of Idiosyncratic Utility	σ	11.6	0.5
Percent College	γ_{Coll}	7.5	0.6
Percent Democrat	γ_{Pol}	17.6	1.7
Constant	γ_0	-308.4	13.2
Number of Panels	γ_{1N}	34.1	1.6
Number of Panels Squared	γ_{2N}	-1.1	0.1

Table 2: Parameter estimates for household utility function. Standard errors calculated via bootstrapping. "Percent Democrat" refers to the percent of voters who voted Democrat in the 2016 presidential election.

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 185 dollars per ton of CO22, based on the mean estimate from Rennert et al. (2022).[39]

5 Estimation Results and Model Fit

5.1 Households

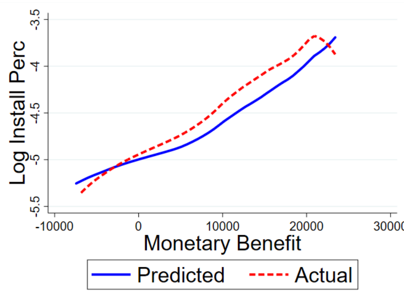
5.1.1 Parameter Estimates

Table 2 displays the estimates of parameters governing the household utility function. The nonpecuniary value of installations is increasing in average local education and in the percent of voters who voted Democrat in the 2016 presidential 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 $\left|\frac{1}{2\times(-1.1)}\right| \approx 0.45$ increase in the optimal number of panels.[40]

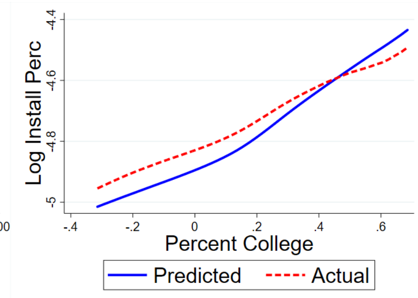
To get a better sense of what the parameter estimates imply for installation probabilities, recall that the partial elasticity of installation probability with respect to monetary benefits is

[39] Rennert et al. (2022) is the same methodology and results on which US EPA based its December 2023 guidance for the social cost of greenhouse gasses. They report estimates in 2020 US Dollars. We convert this to 2014 dollars for consistency with the rest of our analysis. 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).

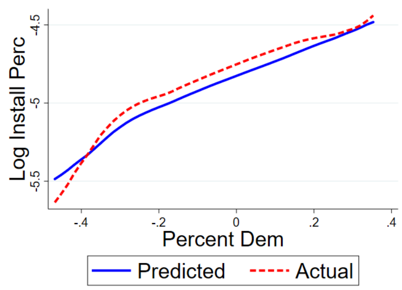
[40] Recall from equation (9) that the optimal number of panels is given by $N_i^\star = \min\left[\bar{N}_i, -\left(\frac{\frac{\partial \mu_{ij}}{\partial N_i}+\gamma_{1N}}{2\gamma_{2N}}\right)\right]$. We estimate $\gamma_{2N} = -1.1$ and increasing the per-panel subsidy by \$1000 increases $\frac{\partial \mu_{ij}}{\partial N_i}$ by 1.



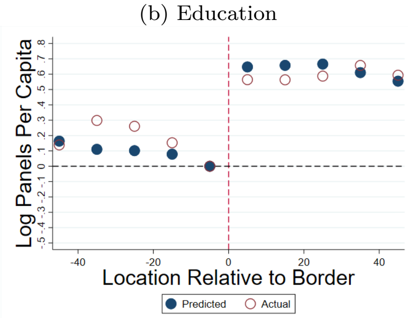
(a) Monetary Benefit



(b) Education



(c) Percent Democrat



(d) Border Discontinuities

Figure 4: Panels (a), (b), and (c) show local nonparametric fit of tract-level log installation per household in the data (red dotted line) and simulations (solid blue line) on (a) the total monetary benefit of installing 15 solar panels, (b) the percent of households with a college degree, and (c) the percent of voters who voted Democrat in the 2016 presidential election. Monetary benefits are measured in 2014 dollars. Panel (d) plots estimated location-bin fixed effects from a regression of log panels per household on state tax rates and location-bin and state-border fixed effects. Positive (negative) values on the X-axis represent households on the side of the border with more (less) generous subsidies. The dots represent estimates of location-bin fixed effects from the data (red hollow dots) and the simulations (blue dots).

approximately equal to $\frac{1}{\sigma}$.[41] Given that we measure monetary values in thousands of dollars, our estimate of $\sigma = 11.6$ implies that a thousand dollar increase in the monetary value of installation leads to approximately a $\frac{1}{11.6} \approx 9$ percent increase in the number of installations.

5.1.2 Model Fit (Installations)

Figure 4 assesses model fit with regard to solar installations. Figure 4a 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.[42] Subfigures 4b and 4c show the relationship between installations and the percentage

[41]Differentiating equation (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.

[42]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.

of households with a college education, and the percent of voters who voted Democrat in the 2016 presidential election. The fit is quite good in both dimensions.

Subfigure 4d examines how log panels per household change as we cross state borders. We first divide households into 10-mile-wide bins in location relative to the border using the procedure described in Section 4.2. We then regress tract-level log panels per household on location-bin fixed effects, controlling for border fixed effects and state tax rates. Finally, we plot the estimated distance-bin fixed effects for both the data (blue dots) and our simulated model (red dots). Positive values on the X-axis represent households on the side of the border with more generous subsidies, and negative values on the X-axis indicate the side with less generous subsidies. Both data series show a similar "jump" as we move from the side of the border with less generous benefits to the side with more generous benefits.

5.1.3 Comparison to Existing Estimates

Gillingham and Tsvetanov (2019) estimate the price elasticity of demand for solar panel installations using an approach that accounts for excess zeroes, unobserved heterogeneity, and the endogeneity of installation prices. Their estimates imply a price elasticity of demand evaluated at the mean installation price equal to -0.65. We simulate a marginal increase in installation prices and calculate the implied price elasticity evaluated at the mean installation price. This yields an estimate of -0.74, close to the elasticity estimated by Gillingham and Tsvetanov (2019).

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 36%, of a similar magnitude to the estimates in Crago and Chernyakhovskiy (2017).[43]

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 $.47 \times \frac{1}{\sigma} \approx 4$ 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.

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

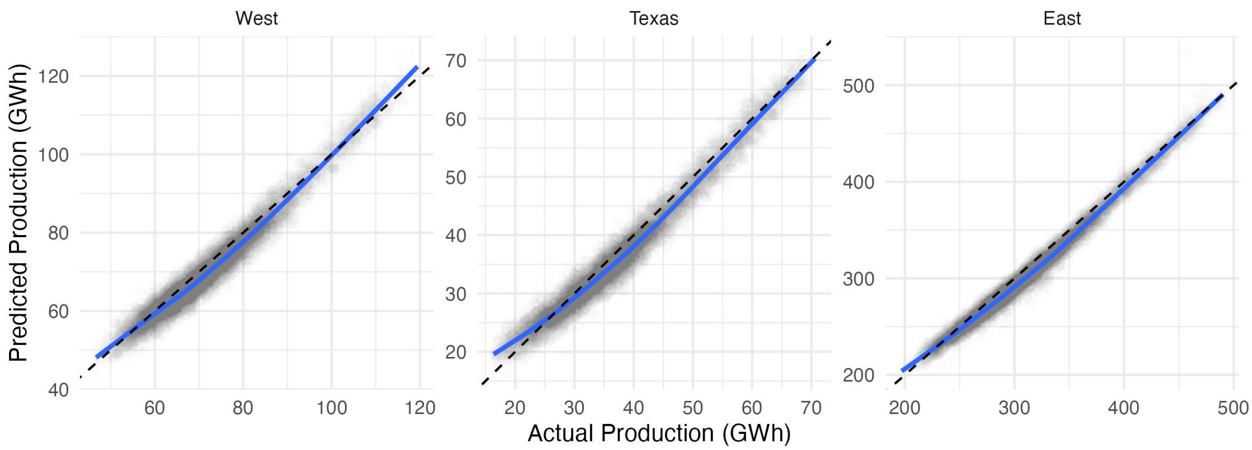


Figure 5: Model fit at the interconnection level. The Y-axis gives the production of dispatchable plants predicted by the model, measured in GWh while the X-axis gives the production in the data. Dots represent an hour of production for each interconnection, smoothed lines show the fit of a generalized additive model.

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 Online Appendix C.6, including out-of-sample predictions using 2020 data. Figure 5 shows predicted electricity production of dispatchable plants against actual electricity production.[44] 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 6. 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. In particular, our model is able to generate the pattern of increasing dispatchable production through the afternoon, the time where solar power generation decreases and electricity demand increases. This is especially true for seasons and interconnections when solar makes up a larger share of electricity production. A region-level breakdown of these plots is available in Online Appendix C.6.

Not only does the model match total production, but it also replicates changes in the

[44]While generally performing very well, we note that these predictions are lower than actual production values by about 2 percent on average. The Tobit model relies on normal and homoskedastic errors to produce consistent estimates (Cameron and Trivedi, 2005, p. 538). As these models are power plant specific, this assumption applies to the errors *within* power plants, some of which may violate the normality assumption. We do not think this issue meaningfully affects our results as the discrepancy between actual and predicted production is small.

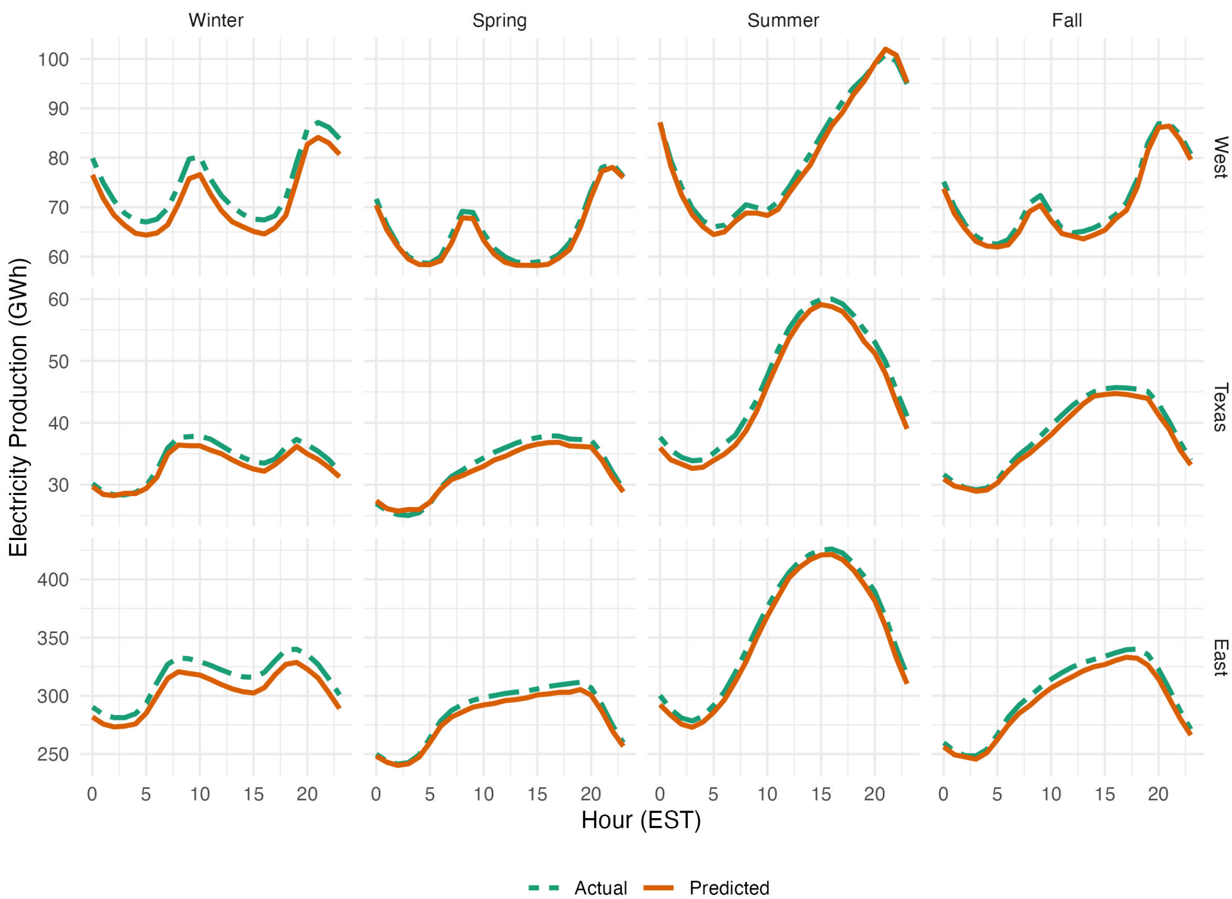


Figure 6: Model fit at the interconnection level by hour and season. Hour reflects Eastern Standard Time (EST). Each panel shows predicted and actual production of dispatchable plants over the course of the average day, for each of the three intersections and across four seasons. The dashed green line gives electricity production in the data while the solid orange line gives predicted production.

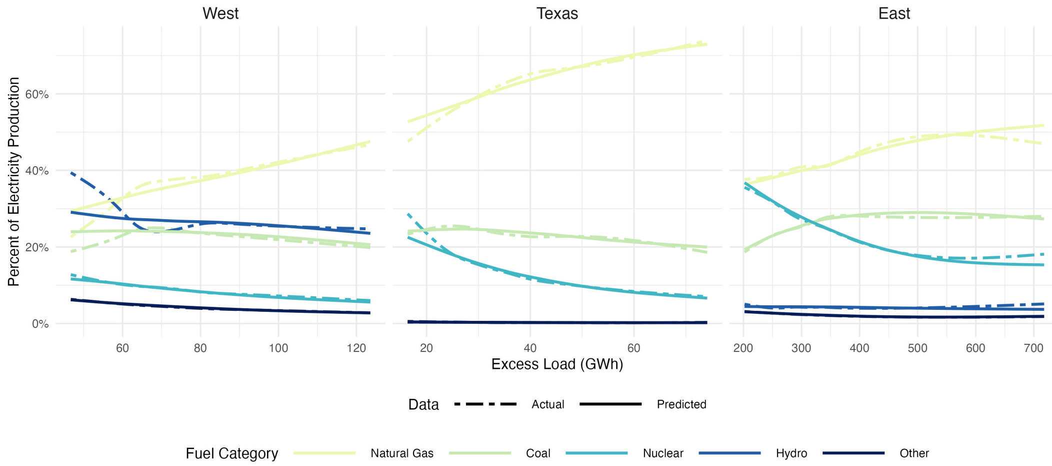


Figure 7: Fuel mix of production by interconnection. The X-axis gives excess load at the interconnection level and the Y-axis gives the percent of electricity production that is produced by each of the fuel types. The dashed lines show the fuel mix in the data while the solid lines show the simulated fuel mix.

fuel mix at varying demand levels, reflecting that plants differ in how balancing authorities dispatch them as a function of excess demand. Figure 7 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, 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's production share declines 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 8 plots simulated marginal damages of energy production in each region within each interconnection as a function of excess load.[45] 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

[45]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.

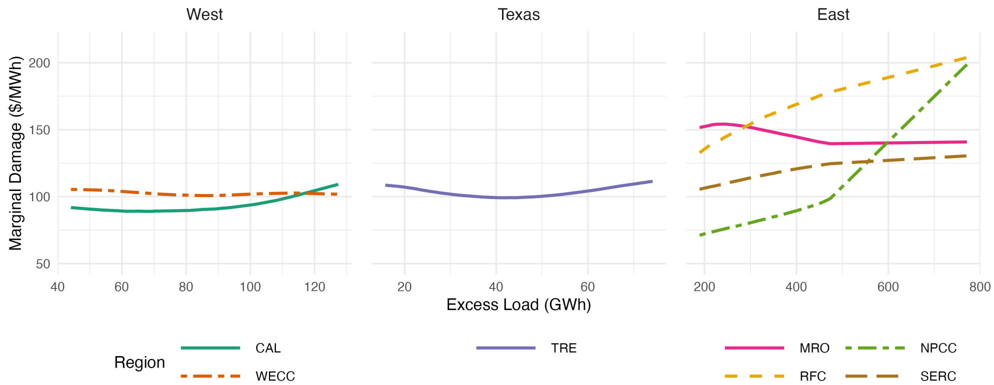


Figure 8: Estimated marginal damage of electricity production by region. The X-axis of each panel varies the total excess load in each of the three interconnections, and the Y-axis gives the simulated marginal damages per MWh of electricity produced in each region. In the Western interconnection, CAL covers most of California and WECC the rest of the western states. TRE is the only region in Texas. In the eastern interconnection, MRO covers the Midwest, NPCC the Northeast, RFC the mid-Atlantic, and SERC the Southeast. See Appendix A.7 for a detailed description of the regions and Figure A5 for service areas of each region. See text for additional details on marginal damage calculations.

from power plants the RFC region, which spans much of the Mid-Atlantic and lower Great Lakes. 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 relatively flat as a function of excess load while marginal damages in several regions in the Eastern interconnection are strongly increasing in excess load. For example, in the NPCC, the region covering the Northeast, marginal damages increase by over 30% between the 10th to 90th percentile of excess load, going from \$67/MWh to \$88/MWh.

6 Counterfactuals and Nationally-Optimal Subsidies

6.1 Welfare-Maximizing Subsidies

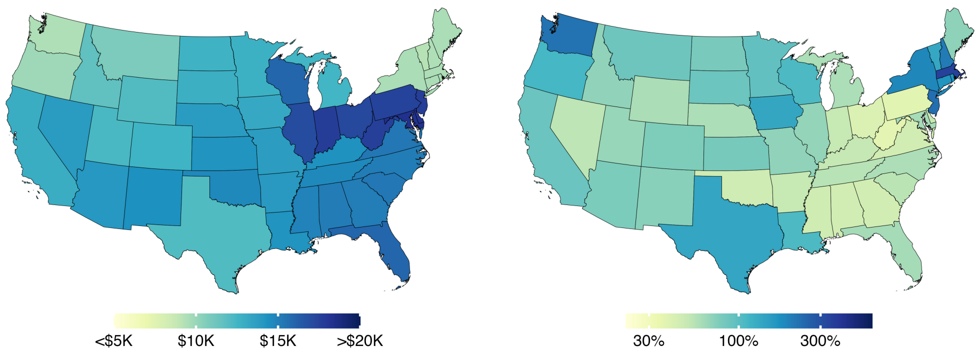
We now use the estimated structural model to quantify the welfare-maximizing solar subsidies characterized by equation (8). We outline the algorithm we use to numerically solve for welfare-maximizing subsidies in Online Appendix B.4. The results are displayed in Figure 9 and in Table 3. In all tables, "Baseline" refers to simulated outcomes under the current system of subsidies.

Figure 9a and the first panel of Table 3 show how total subsidy generosity varies across states under the nationally-optimal system of subsidies.[46] We measure subsidy generosity as the present discounted value of subsidies an average household in each state would receive if

[46] In Online Appendix C.7, we show how the government should optimally allocate subsidies across the three subsidy types: cost-based subsidies, per-panel subsidies, and production-based (per-kWh) subsidies.

	(1)	(2)	(3)	(4)	(5)
		State-Specific Subsidies		Tract-Specific Subsidies	
	Baseline	Welfare Max	Damage Min	Welfare Max	Damage Min
I. Average Subsidy ($Thousands)					
Midwest	10.0	15.4	15.9	15.3	15.8
Northeast	17.1	13.1	11.2	13.1	11.1
South	10.6	14.6	14.5	14.6	14.4
West	11.7	12.6	10.5	12.6	10.3
II. Installations per 1000HHs					
Midwest	5.4	8.1	8.9	8.1	8.9
Northeast	20.3	11.3	11.0	11.3	10.9
South	6.6	9.0	9.3	9.0	9.3
West	11.9	13.1	11.3	13.1	11.2
National	9.9	10.3	10.0	10.3	10.0
III. Annual Damages Offset ($Millions)					
CO2e	193.4	216.9	219.0	217.1	219.4
NOx	12.0	13.3	13.3	13.3	13.3
PM2.5	33.2	36.8	34.7	36.8	34.6
SO2	23.6	25.5	28.5	25.7	28.9
Total	262.3	292.6	295.4	292.9	296.1

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. All monetary values are measured in 2014 dollars. See text for details on each simulation.



(a) National optimal subsidies. (b) Baseline installations as a percent of optimal.

Figure 9: State-level nationally-optimal subsidies and misallocation for welfare-maximizing reforms. Panel (a) gives the nationally-optimal state subsidies. Subsidies are measured as the present discounted value associated with a 15-panel installation in 2014 dollars, averaged across all households in the state. Panel B shows state-level installations under the current system as a percentage of installations under the nationally-optimal system. These results are shown in table form in Online Appendix C.9.

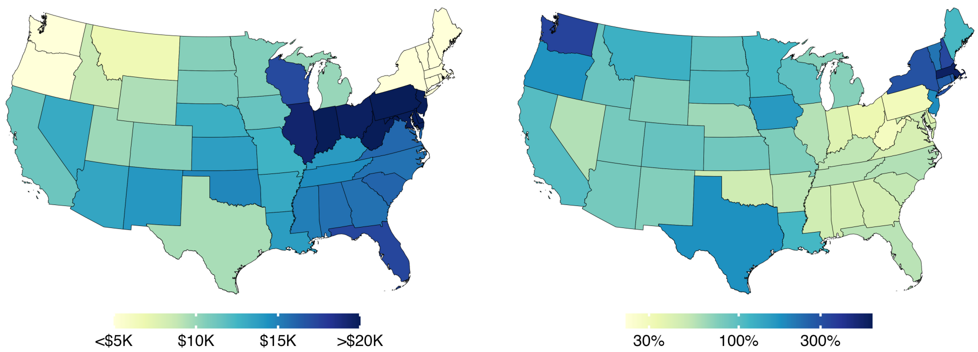
they installed a 15-panel system. Several states in the Northeast and Northwest, areas with relatively environmentally friendly power plants and little sunlight, have the lowest optimal subsidies, at under 10 thousand dollars in present value. On the other end of the spectrum, eight states have optimal subsidies valued at over 17 thousand dollars. In West Virginia, one of these states, current subsidy levels are some of the least generous in the country at under 6 thousand dollars. More generally, nationally-optimal subsidies are highest in the Mid-Atlantic and lower Great Lakes and are lowest in the Northwest and Northeast.

Figure 9b 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.[47] Current installations in the Midwest and South are roughly 30% lower than under nationally-optimal subsidies, while installations in the Northeast are 80% higher than the optimal level. 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 nationally-optimal scheme would increase annual damages offset by rooftop solar from $262.3 million to $292.6 million, equal to a 11.5% increase in the aggregate environmental benefits of solar panels.[48] A decrease in CO2 equivalent

[47]We find that almost all of the adjustment comes via the extensive margin, rather than the intensive margin (number of panels per installation). In Online Appendix C.8, we show how the average installation size changes across counterfactuals.

[48]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 directly allocate panels across states subject to local capacity constraints. Here we consider a government which can only influence installations through subsidies. Second, we utilize emissions data from 2019 rather than 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



(a) Nationally-optimal subsidies (b) Baseline installations as a percent of optimal.

Figure 10: State-level nationally-optimal subsidies and misallocation for damage-minimizing reforms. Panel (a) gives the nationally-optimal state subsidies. Subsidies are measured as the present discounted value associated with a 15-panel installation in 2014 dollars, averaged across all households in the state. The color scale censors subsidy levels below \$5K and above \$20K. Panel B shows state-level installations under the current system as a percentage of installations under the nationally-optimal system. These results are shown in table form in Online Appendix C.9.

emissions drives most of the environmental gains, with relatively minor effects on damages from other pollutants.

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, $D\,(\text{ELoad})$, subject to the government budget constraint. We formalize the government's problem and present the first-order conditions in Online Appendix B.5.

Figure 10 and the third columns of Tables A8 and Table 3 show the results. Like the welfare-maximizing policies, the damage-minimizing policies are most generous in the Mid-Atlantic and lower Great Lakes, and are least generous in the Northwest and Northeast. However, the variation across states in subsidy generosity is greater than under the welfare-maximizing subsidies: optimal damage-minimizing subsidies range from under 4 thousand dollars to over 21 thousand dollars. The reallocation of solar panels induced by the damage-minimizing subsidies would lead to a 12.6% 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

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.

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.[49]

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 nationally-optimal system of state-level subsidies can capture most of the gains of more geographically granular subsidies.

6.4 Unconstrained Reforms

Our previous counterfactuals have focused on budget-neutral reforms. Here we analyze the case where the government does not face an externally set budget constraint and maximizes utility less environmental damages and government cost. In this case, the government's problem is to maximize

$$\underbrace{\int_i V_i di}_{\text{Utility}} - \underbrace{D\left(\text{ELoad}\right)}_{\text{Damages}} - \lambda \underbrace{\sum_j \int_{i \in I_j} s_{ij} m_i^{\star} di}_{\text{Government Cost}}, \tag{14}$$

where λ represents the marginal cost of public funds. We present results over a range of λ's from 1 to 1.5—for $\lambda = 1$, the government weights fiscal costs and household utility equally. We can think of this as a setting in which there are no distortionary costs of raising public funds, e.g. the government has access to lump-sum taxation.[50] We present the first-order conditions of the government's problem in Online Appendix C.10.

The results are summarized in Table 4. The first two columns show results under the current subsidy system and unconstrained nationally-optimal subsidies when we use a marginal cost of funds (λ) equal to 1. Optimal subsidies slightly exceed current subsidies in the Midwest and South, but fall below current levels in the Northeast and West. Table A12 in

[49] 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 nationally-optimal system of subsidies when the planner can use both panel-based and production-based subsidies. Therefore, we set per-panel subsidies to 0 in this exercise.

[50] Jacobs (2018) shows that the marginal cost of funds is also equal to one in a Mirrleesian framework with the optimal tax system. A value of 1.3 is a commonly-used "middle of the road" estimate of the marginal cost of public funds in the literature (Ballard, Shoven, and Whalley, 1985; Newhouse, 1992; Poterba, 1996; Olken, 2007; Finkelstein and McKnight, 2008) while 1.5 is considered a "conservative" estimate (Heckman et al., 2010; Finkelstein and Hendren, 2020).

	(1)	(2)	(3)	(4)
		Unconstrained Optimal		
	Baseline	$\lambda = 1$	$\lambda = 1.25$	$\lambda = 1.5$
I. Average Subsidy ($Thousands)				
Midwest	10.0	11.9	7.2	4.1
Northeast	17.1	9.9	5.6	2.7
South	10.6	11.2	6.7	3.6
West	11.7	9.5	5.3	2.5
II. Installations per 1000HHs				
Midwest	5.4	6.0	4.1	3.1
Northeast	20.3	8.6	5.9	4.6
South	6.6	6.7	4.6	3.5
West	11.9	10.1	7.0	5.5
National	9.9	7.8	5.4	4.2
III. Annual Damages Offset ($Millions)				
Total	262.3	218.6	147.4	113.4
IV. Annuitized Total Subsidies Paid ($Millions)				
National	380.8	218.5	86.2	34.7

Table 4: Unconstrained Nationally-Optimal Subsidies. 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. The third panel gives the annual environmental benefits generated by residential solar panels. The final panel gives the total government spending on subsidies converted into an annuity value. All monetary values are measured in 2014 dollars.

Online Appendix C.10 shows the state-level optimal subsides. These optimal subsidies result in fewer installations nationally, with Panel II showing that installations under the optimal subsidies are 80% of current levels nationally.

Panels III and IV show the annual environmental benefits of rooftop solar and the annuitized government spending on subsidies. Switching to unconstrained nationally-optimal subsidies decreases environmental benefits by 43 million dollars annually. However, the accompanying 162 million dollar decrease in fiscal costs dwarfs this decrease in environmental benefits. The optimal subsidy scheme achieves 83% of the environmental benefits at 57% the current cost.

To account for the additional costs of raising funds for solar subsidies from the use of distortionary taxes, we now calculate the optimal unconstrained subsidy with alternative values of λ, the marginal costs of public funds. In Columns 3 and 4 of Table 4, we recalculate the optimal subsidies with $\lambda = 1.25$ and with $\lambda = 1.5$, respectively. With these larger values of the marginal cost of public funds, the optimal amount of spending on solar subsidies decreases, as raising money for subsidies entails an efficiency cost. With $\lambda = 1.25$, optimal spending on subsidies is only 23% of current levels, and with $\lambda = 1.5$, optimal spending is less than one-tenth of current levels.

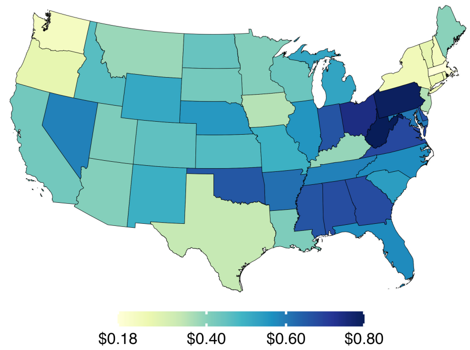


Figure 11: Damages offset per additional dollar of government funds associated with marginal increases in production-based subsidies, s_j^{kWh}, around the current system of subsidies. All monetary values are measured in 2014 dollars.

As discussed in the introduction, residential solar subsidies may be associated with additional external benefits that we do not capture here. However, our results suggest that these additional benefits would have to be quite large to justify the current spending levels on these subsidies nationally.

6.5 Marginal Subsidy Increases

Relative to 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. Specifically, we first simulate the model 1) under the current system of subsidies and 2) with marginally more generous subsidies of a given type in a given state. We calculate the damages offset per dollar of this particular subsidy as the difference in damages between the two simulations divided by the difference in the fiscal cost. We repeat this process for each subsidy type in each state.

Figure 11 shows the marginal damages offset per dollar for production-based subsidies.[51] There are large differences in damages offset across states. For example, a small subsidy increase in Washington only leads to 22 cents less environmental damages per dollar 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 80 cents. Subsidy increases in Ohio and Pennsylvania are also associated with damages offset per dollar of over 70 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

[51]Within each state, there are only small differences in the damages offset per dollar across the three subsidy types.

would lead to decreases in environmental damages of $80 - 22 = 58$ 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.

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 Online Appendix D.1, we assess the sensitivity of our main 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 cost-neutral 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 Alternative Discounting of Future Subsidy Payments

In our baseline model, we assumed that households value production subsidies at the present discount value of subsidy payments, calculated using the market interest rate. However, De Groote and Verboven (2019) find that households use a much higher implicit interest rate when evaluating the future subsidy payments associated with solar panel installations.

In Online Appendix D.2, we reestimate our model and recalculate the optimal subsidies when households calculate the net present value of future subsidy payments using an implicit discount rate of 15%, based on the estimates from De Groote and Verboven (2019). We have two main takeaways from this extension. First, the change in household discounting of future subsidy payments has important implications for the *type* of subsidy the planner chooses. When households discount future subsidies more heavily than upfront subsidies, the social planner relies less on production subsidies, as they are valued less than upfront subsidies.[52] Therefore, in this extension, the social planner relies mostly on investment subsidies, whereas the planner in our baseline results uses production subsidies most heavily. Our results echo those from De Groote and Verboven (2019), who find that upfront investment subsidies are more cost-effective than production subsidies at inducing installations. Second, the optimal

[52] We do not allow the government to set negative production subsidies. Without this constraint, the government will heavily tax solar production.

geographic distribution of subsidies and the optimal overall level of subsidy spending are similar to our baseline results. We conclude that this alternative assumption about household discounting does not change our main results about spatial misallocation of panels nor about optimal levels of spending across space.

7.3 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 Online Appendix D.3, 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 central generation 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.

The takeaways are qualitatively the same as our main results. As expected, the environmental benefits of solar panels increase. As a result, the total government spending under the optimal unconstrained subsidies is higher than in the baseline case but still less than under current policies. The welfare-maximizing and damage-minimizing cost-neutral reforms generate increases in aggregate environmental benefits of 11.4% and 12.3%, respectively.

7.4 Transmission Constraints

In times when transmission lines are under heavy use, grid operators will limit the extent to which electricity is moved over long distances to prevent transmission lines from overheating. Our baseline model does not account for these transmission constraints because we assume that a given power plant's production level can always depend on excess load in other regions within the plant's interconnection. If transmission constraints are binding, however, the plant's power production is unlikely to depend on excess load in other regions because grid operators will avoid transmitting electricity across regions.

We now consider a model extension that includes a stylized model of potentially binding transmission constraints. We assume that if the excess load in a given region R exceeds a certain threshold $\overline{\mathrm{ELoad}}_R$, then the policy functions of power plants will depend only on the excess load in their own region and not on excess load in other regions. Thus, we can express the latent function for power plant k in region R as:

$$f_k\left(\text{ELoad}_t, \varepsilon_{kt}\right) = \begin{cases} \psi_k^0 + \sum_{R' \in \mathbf{R}_k} \left(\psi_{R'k}^1 \text{ELoad}_{R't} + \psi_{R'k}^2 \text{ELoad}_{R't}^2\right) + \varepsilon_{kt} & \text{if } \text{ELoad}_{Rt} \leq \overline{\text{ELoad}}_R \\ \psi_k^{0,\text{Cons}} + \psi_k^{1,\text{Cons}} \text{ELoad}_{Rt} + \psi_k^{2,\text{Cons}} \text{ELoad}_{Rt}^2 + \varepsilon_{kt} & \text{if } \text{ELoad}_{Rt} > \overline{\text{ELoad}}_R \end{cases}. \quad (15)$$

In practice, we set $\overline{\text{ELoad}}_R$ as the 75th percentile of excess load in a given region in our data. We re-estimate the power plant policy functions via maximum likelihood and recalculate the optimal policies with the updated policy functions. We present the results in Appendix D.4. As in our baseline specification, we find that optimal unconstrained subsidies are lower than current levels in the Northeast and the West, but are similar to or slightly above current levels in the Midwest and South. We find that current spending on subsidies exceeds the optimal level by about 45% in this specification.

7.5 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 Online Appendix D.5, 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.

We find that the storage technology itself generates considerable environmental benefits. However, storage technology does not substantially change the environmental benefits of solar panels across locations. Therefore, we find that adding storage technology does not qualitatively change the optimal cost-neutral or unconstrained reforms, the distribution of installations under the optimal subsidies, or the environmental benefits of switching to optimal subsidies. See Butters, Dorsey, and Gowrisankaran (2021) or Holland, Mansur, and Yates (2022) for a detailed treatment of storage technology.

7.6 Cleaner Electricity 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 electricity production would affect our results. First, we find the optimal subsidies under expanded production from utility-scale solar and wind based on three scenarios of projected renewable expansion by 2030 from the EIA (Nalley and LaRose, 2022). 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.

We present the results in Online Appendix D.6. Intuitively, we find that if electricity production becomes cleaner, the environmental damages offset by solar panels will decrease. Therefore, the optimal unconstrained subsidies for residential solar with cleaner electricity production are lower than in our baseline case, suggesting optimal unconstrained subsidies will be even lower in the future if electricity production continues to become cleaner. However, there remains considerable heterogeneity in the marginal damage of electricity production across space. Thus, we still find substantial benefits in switching from the current subsidies to optimal cost-neutral subsidies.

7.7 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). Our proposed nationally-optimal subsidies will likely be progressive relative to the current subsidies since switching from current to optimal generally involves decreasing subsidies in high-income states such as Massachusetts and increasing subsidies in low-income states such as West Virginia. For this same reason, switching to optimal subsidies will likely improve the distribution of damages caused by electricity generation. Similarly, Dauwalter and Harris (2023) find that shifting solar capacity to locations where the environmental

	Dependent Variable: Log Installations per Capita			
	(1)	(2)	(3)	(4)
Historically-Adjusted Subsidies ($1000s)	0.0592***	0.0604***	0.0597***	0.0707***
	(0.00588)	(0.00472)	(0.00549)	(0.00882)
Income Tax	NO	YES	YES	YES
Sales Tax	NO	NO	YES	YES
Property Tax	NO	NO	NO	YES
Distance Bandwidth	10	10	10	10
Observations	6,052	6,052	6,052	6,052

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

Table 5: Regression of log installations per capita on "historically-adjusted" measures of subsidy generosity from Sexton et al. (2021). Sample is limited to tracts within 10 miles of state borders. Historically-adjusted subsidy generosity measures the net present value of subsidies a household would expect to receive given subsidies in place in the year of installation, measured in thousands of 2014 dollars.

benefits are greatest would lead to 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.[53]

7.8 Installation Elasticities with Historical Subsidy Measures

As described earlier, our model of solar installation captures the long-run installation behavior of households facing a given menu of subsidies. One potential issue is that we estimate the model using solar subsidies from 2017 when, in fact, many of the installations observed in our data were made before 2017 by households facing pre-2017 subsidy levels.

To get a sense of how accounting for pre-2017 subsidies would affect our estimates of the household installation model, we perform a simple exercise using a measure of "historically-adjusted" subsidy generosity from Sexton et al. (2021). To calculate this measure, the authors first calculate the net present value of subsidies a household in a given state would expect to receive, given the subsidies in place from each year from 2000 to 2017. They then take the average over these yearly measures, weighing them by the number of installations in each year.[54] This is a measure of the net present value of subsidies from a solar installation,

[53] Colas and Reynier (2024) solve for the optimal income-contingent subsidies for residential solar panels.

[54] To generate the subsidy measures reported in Sexton et al. (2021), the authors then divide the net present value of subsidies by the total kWh of electricity an installation is expected to produce, such that the reported

adjusting for the fact that households have faced different subsidies based at the time when they installed.

We then replicate the border discontinuity regressions from Table 1 using this historically adjusted measure of the net present value of subsidies. Specifically, we regress tract-level log installations per capita on this historically adjusted measure of the net present value of subsidies, limiting our sample to tracts within 10 miles of state borders. Table 5 shows the results. Across all specifications, we estimate that a 1000 dollar increase in the historically-adjusted net present value of subsidies is associated with a 5.9% to 7.1% increase in installations per capita, similar to the results we found in Table 1, when we used our baseline subsidy measures. These results suggest that not accounting for the profile of subsidies households have faced over time may not introduce significant bias when estimating the long-run responsiveness of installation to subsidies.

8 Conclusion

We have used a structural model of solar panel demand and electricity production to calculate the nationally-optimal system of subsidies for residential 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 a spatial misallocation of panels, and nationally spending on subsidies is too high. However, our results do not necessarily imply that the US should lower funding for renewable energy programs in general, rather that government funds spent on subsidies for residential solar subsidies would be better spent on other programs. These alternative programs could include other investments in renewable energy, such as subsidies for utility-scale solar or wind power, both of which provide energy at lower cost than residential solar (Lazard, 2023).

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 disincentivize 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 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.

measure represents the subsidy received per kWh. We do not perform this conversion so that our measure represents the total net present value of subsidies for a given installation, measured in thousands of dollars. This makes our estimates here easier to compare to our regressions in Section 4.2.

References

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.*

Ballard, Charles L, John B Shoven, and John Whalley. 1985. "General equilibrium computations of the marginal welfare costs of taxes in the United States." *The American Economic Review*, 75(1): 128–138.

Barbose, Galen, and Naïm Darghouth. 2019. "Tracking the Sun: Pricing and Design Trends for Distributed Photovoltaic Systems in the United States - 2019 Edition."

Batini, Nicoletta, Mario Di Serio, Matteo Fragetta, Giovanni Melina, and Anthony Waldron. 2022. "Building back better: How big are green spending multipliers?" *Ecological Economics*, 193: 107305.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. "A unified framework for measuring preferences for schools and neighborhoods." *Journal of Political Economy*, 115(4): 588–638.

Benthem, Arthur van, Kenneth Gillingham, and James Sweeney. 2008. "Learning-by-doing and the optimal solar policy in California." *The Energy Journal*, 29(3): 131–152.

Black, Sandra E. 1999. "Do better schools matter? Parental valuation of elementary education." *The Quarterly Journal of Economics*, 114(2): 577–599.

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

Bollinger, Bryan, and Kenneth Gillingham. 2019. "Learning-by-doing in solar photovoltaic installations." *Available at SSRN 2342406.*

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

Borenstein, Severin. 2012. "The private and public economics of renewable electricity generation." *Journal of Economic Perspectives*, 26(1): 67–92.

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

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.

Brown, Patrick R, and Francis M O'Sullivan. 2020. "Spatial and temporal variation in the value of solar power across United States electricity markets." *Renewable and Sustainable Energy Reviews*, 121: 109594.

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.

Cameron, A Colin, and Pravin K Trivedi. 2005. *Microeconometrics: methods and applications.* Cambridge university press.

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, and Emmett Reynier. 2024. "Means-Tested Solar Subsidies." *Working Paper.*

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.

Dorsey, Jackson. 2024. "Solar Market Frictions: The Role of Platforms and Policies." *Working Paper.*

Eichner, Thomas, and Marco Runkel. 2014. "Subsidizing renewable energy under capital mobility." *Journal of Public Economics*, 117: 50–59.

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.

Finkelstein, Amy, and Nathaniel Hendren. 2020. "Welfare analysis meets causal inference." *Journal of Economic Perspectives*, 34(4): 146–167.

Finkelstein, Amy, and Robin McKnight. 2008. "What did Medicare do? The initial impact of Medicare on mortality and out of pocket medical spending." *Journal of public economics*, 92(7): 1644–1668.

Gerarden, Todd D. 2023. "Demanding innovation: The impact of consumer subsidies on solar panel production costs." *Management Science*, 69(12): 7799–7820.

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

Gillingham, Kenneth, Hao Deng, Ryan Wiser, Naim Darghouth, Gregory Nemet, Galen Barbose, Varun Rai, and Changgui Dong. 2016. "Deconstructing solar photovoltaic pricing: the role of market structure, technology, and policy." *The Energy Journal*, 37(3): 231–250.

Goulder, Lawrence H, and Koshy Mathai. 2010. "Optimal CO2 abatement in the presence of induced technological change." In *Technological change and the environment.* 210–250. Routledge.

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.

Hasna, Zeina. 2021. "The grass is actually greener on the other side: Evidence on green multipliers from the united states." *Job Market Paper.*

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).

Heckman, James J, Seong Hyeok Moon, Rodrigo Pinto, Peter A Savelyev, and Adam Yavitz. 2010. "The rate of return to the HighScope Perry Preschool Program." *Journal of public Economics*, 94(1-2): 114–128.

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. 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.

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.

Ito, Koichiro. 2014. "Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing." *American Economic Review*, 104(2): 537–563.

Jacobs, Bas. 2018. "The marginal cost of public funds is one at the optimal tax system." *International Tax and Public Finance*, 25: 883–912.

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

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.

Lazard. 2023. "2023 Levelized Cost Of Energy+." Lazard.

Lee, David S, and Thomas Lemieux. 2010. "Regression discontinuity designs in economics." *Journal of Economic Literature*, 48(2): 281–355.

Miller, Gregory, Gailin Pease, Milo Knowles, and Wenbo Shi. 2022. "Open Grid Emissions Initiative (v0.3.0)."

Millstein, Dev, Ryan Wiser, Mark Bolinger, and Galen Barbose. 2017. "The climate and air-quality benefits of wind and solar power in the United States." *Nature Energy*, 2(9): 1–10.

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.

Newhouse, Joseph P. 1992. "Medical care costs: how much welfare loss?" *Journal of Economic perspectives*, 6(3): 3–21.

Olken, Benjamin A. 2007. "Monitoring corruption: evidence from a field experiment in Indonesia." *Journal of political Economy*, 115(2): 200–249.

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

Pless, Jacquelyn, and Arthur A Van Benthem. 2019. "Pass-through as a test for market power: An application to solar subsidies." *American Economic Journal: Applied Economics*, 11(4): 367–401.

Pless, Jacquelyn, Ria Langheim, Christina Machak, Henar Hellow, and Ben Sigrin. 2017. "The price-concentration relationship in early residential solar third-party markets." National Renewable Energy Lab.(NREL), Golden, CO (United States).

Popp, David, Francesco Vona, Giovanni Marin, and Ziqiao Chen. 2020. "The employment impact of green fiscal push: evidence from the American Recovery Act." National Bureau of Economic Research.

Poterba, James M. 1996. "Government intervention in the markets for education and health care: how and why?" In *Individual and social responsibility: Child care, education, medical care, and long-term care in America*. 277–308. University of Chicago Press.

Reguant, Mar. 2019. "The efficiency and sectoral distributional impacts of large-scale renewable energy policies." *Journal of the Association of Environmental and Resource Economists*, 6(S1): S129–S168.

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.

Rubin, Edward, and Maximilian Auffhammer. 2023. "Quantifying Heterogeneity in the Price Elasticity of Residential Natural Gas."

Saez, Emmanuel, and Stefanie Stantcheva. 2016. "Generalized social marginal welfare weights for optimal tax theory." *American Economic Review*, 106(1): 24–45.

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

Siler-Evans, Kyle, Inês Lima Azevedo, M Granger Morgan, and Jay Apt. 2013. "Regional variations in the health, environmental, and climate benefits of wind and solar generation." *Proceedings of the National Academy of Sciences*, 110(29): 11768–11773.

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

Tibebu, Tiruwork B, Eric Hittinger, Qing Miao, and Eric Williams. 2021. "What is the optimal subsidy for residential solar?" *Energy Policy*, 155: 112326.

Walczak, Jared, and Scott Drenkard. 2017. *State and Local Sales Tax Rates in 2017.* Washington: Tax Foundation.

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.

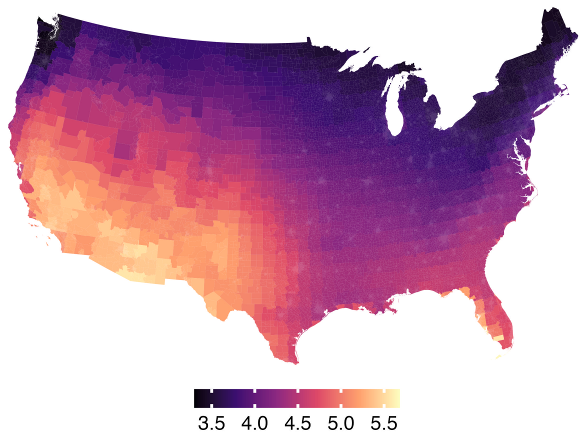


Figure A1: Daily solar radiation $(kWh/m^2/day)$ by census tract from Deepsolar.

A Data Appendix: For Online Publication

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 M_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 the 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 Online 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 2014 and 2018 that are not missing price or the total number of panels, which leaves us with over 720 thousand observations. Additionally, we censor installation costs at the 0.5th and 99.5th percentiles and convert them into 2014 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 county (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),

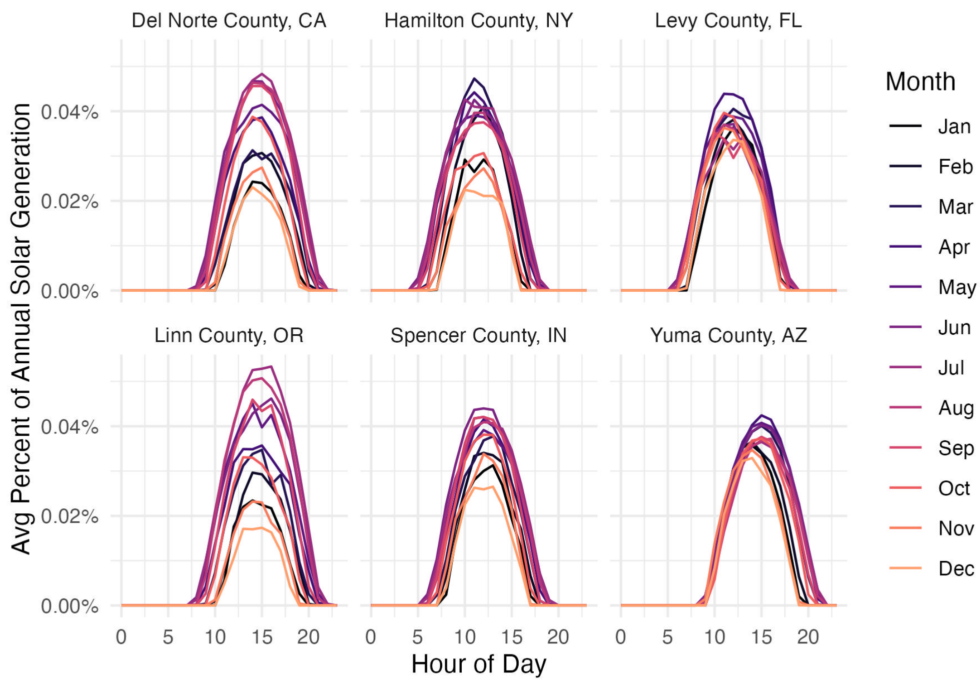


Figure A2: Hourly electricity generation for a standard solar panel for six example counties.

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 county. We use weather data from Sengupta et al. (2018) for a county's typical meteorological year.

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

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

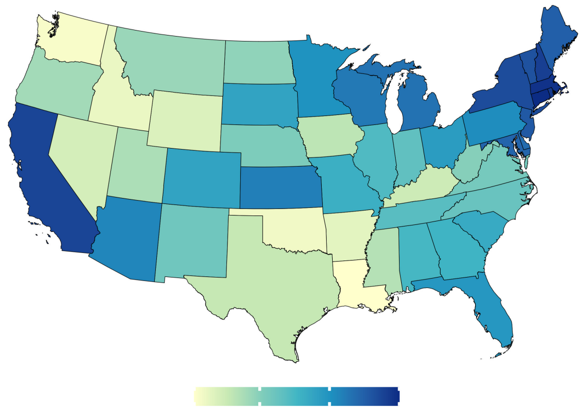
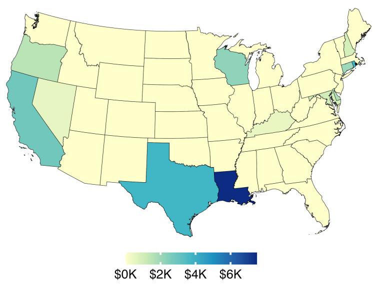
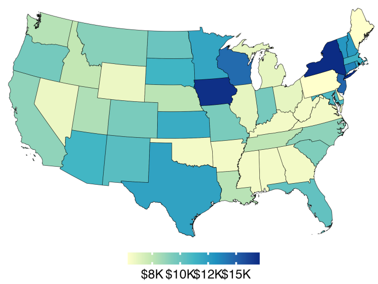


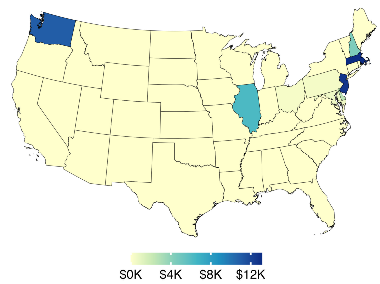
Figure A3: State electricity prices ($\$/kWh$)



(a) Panel-based subsidies



(b) Cost-based subsidies



(c) kWh-based subsidies

Figure A4: Expected subsidy for a 15-panel system by subsidy type in 2014 dollars.

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 exemptions, and the net present value of property tax exemptions. We translate per-KW rebates to s_j^{Panel} by assuming a constant 0.25 KW per panel. Maryland has a fixed rebate of $1000 per system. We translate this into a per-panel subsidy by dividing this amount by the average number of panels in an installation (15). Many states place a cap on the maximum amount of a type of 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: per-panel, cost, and kWh. Most of the current subsidies take the form of cost-based subsidies, while few states offer kWh and panel-based subsides.

A.7 Power Plants

The following describes how we use the Open Grid Emissions (OGE) data.[55] These data have several advantages over the commonly used raw electricity data from the EIA and EPA, which we describe below.

[55]OGE is a relatively new dataset and under active development. We use v0.3.0 in this analysis.

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 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.[56]

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 9,167 plants with hourly production and emissions in the 2019 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 4,312 non-dispatchable plants (wind and solar), 193 plants with zero or negative reported net electricity generation, and 64 plants with no variation in net electricity generation, we have 4,598 power plants—giving us nearly 40 million plant-hour observations.

[56] 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.

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 only applies to less than 10% of total electricity production. We censor PM 2.5 emissions rates at the 95th percentile for each fuel category. A few power plants have emissions rates orders of magnitude larger than is reasonable—these are small plants that do not directly report their hourly emissions or generation in the EPA CAMD data, and thus must be imputed by OGE.[57] For these plants, we set emissions rates equal to the median for their fuel type.

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

[57]These plants' EIA ID's are 2221, 2528, 2503, 50626, and 50931.

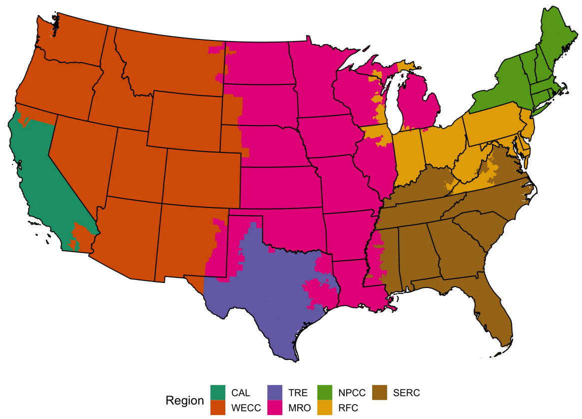


Figure A5: Map of region service areas

Table A1: Summary statistics on power plants by region

		Net Generation		Emissions (lb/MWh)		
Region	Number of plants	Total (TWH)	% Fossil Fuel	NOx	SO2	CO2e
CAL	642	1,623	58.9	0.66	0.07	663
MRO	1,168	8,116	79.7	0.96	1.03	1,343
NPCC	725	2,291	46.1	0.41	0.06	540
RFC	645	7,831	61.8	0.50	0.51	861
SERC	572	9,396	67.4	0.49	0.35	918
TRE	179	3,348	87.2	0.64	0.70	1,069
WECC	667	4,776	61.5	0.75	0.36	1,011
Total	**4,598**	**37,382**	**68.2**	**0.64**	**0.53**	**990**

Table A2: Summary statistics on power plants by fuel category

			Emissions (lb/MWh)		
Fuel	Number of plants	Net generation (TWH)	NOx	SO2	CO2e
Biomass	577	542	5.54	1.12	1,569
Coal	258	9,457	1.44	1.95	2,221
Natural Gas	1,685	16,027	0.41	0.03	906
Nuclear	60	8,136	0.00	0.00	5
Petroleum	517	25	3.98	4.98	1,907
Geothermal	60	155	0.00	0.35	137
Hydro	1,332	2,854	0.00	0.00	0
Other	43	49	0.08	0.00	105
Waste	66	135	5.83	1.32	3,722
Total	**4,598**	**37,382**	**0.64**	**0.53**	**990**

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.[58] 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.

B Theory and Quantitative Appendix: For Online Publication

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.[59] In the states where net metering is not offered, households 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, 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

$$c + \underbrace{p_j \left(e - m_i N_i A_i^{home}(\cdot)\right)}_{\text{Cost of electricity}} + m_i \underbrace{\left[\left(1 - s_j^{\text{Cost}}\right) p_j^{\text{Ins}}(N_i)\right]}_{\text{Net cost of installation}} =$$
$$y_i + m_i \left(\underbrace{N_i A_i s_j^{\text{kWh}}}_{\text{kWh Subsidy}} + \underbrace{N_i s_j^{\text{Panel}}}_{\text{Per-Panel Subsidy}} + \underbrace{p_j^{sale} N_i A_i^{grid}(\cdot)}_{\text{Electricity sold to grid}} \right). \quad (16)$$

In estimating and simulating the model, we assume that the household's optimal elec-

[58]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.

[59]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.

tricity 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

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

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 household's 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,

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

We calibrate C such that a household with the average number of panels sells 30% of their electricity back to the grid.[60]

B.2 Maximum Likelihood Estimation of Power Plant Policy Functions

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

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

$$
\begin{aligned}
\log \mathcal{L}_{kt}\left(\text{ELoad}_t|\psi_k,\sigma_k^2\right) = & \mathbb{1}\left(y_{kt}^o = 0\right) \times \log\left(\Phi\left(\frac{\hat{f}_k\left(\text{ELoad}_t|\psi_k\right)}{\sigma_k}\right)\right) + \\
& \mathbb{1}\left(y_{kt}^o \in (0,\bar{y}_k)\right) \times \log\left(\frac{1}{\sigma_k}\phi\left(\frac{y_{kt}^o - \hat{f}_k\left(\text{ELoad}_t|\psi_k\right)}{\sigma_k}\right)\right) + \\
& \mathbb{1}\left(y_{kt}^o \geq \bar{y}_k\right) \times \log\left(1-\Phi\left(\frac{\bar{y}_k - \hat{f}_k\left(\text{ELoad}_t|\psi_k\right)}{\sigma_k}\right)\right),
\end{aligned} \tag{17}
$$

where $\mathbb{1}(\cdot)$ represents an indicator functions which turns on if y_{kt}^o 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

$$
W = \int_i V_i di - D\left(\text{ELoad}\right) - \lambda\left(\sum_j \int_{i \in I_j} s_{ij} m_i^\star di - G\right), \tag{18}
$$

where $D\left(\text{ELoad}\right)$ is total environmental damages, $s_{ij} = s_j^{\text{Panel}} N_i^\star + s_j^{\text{kWh}} A_i N_i^\star + s_j^{\text{Cost}} p_j^{\text{Ins}}\left(N_i^\star\right)$ is the total subsidy paid to household i conditional on installation, and G is the maximum amount the government can spend on subsidies.

The nationally-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 $\theta \in \{\text{kWh}, \text{Panel}, \text{Cost}\}$ in state j yields

$$
\begin{aligned}
\frac{\partial W}{\partial s_j^\theta} = & \int_i \frac{\partial V_i}{\partial s_j^\theta} di + \int_i \sum_{t=0}^{T} \frac{A_{it}}{(1+r)^t}\left|\frac{\partial D_t\left(\text{ELoad}_t^{SB}\right)}{\partial E_{Rt}^{\text{Solar}}}\right|\left(\overrightarrow{m_i}^\theta N_i^\star + m_i^\star \frac{\partial N_i}{\partial s_j^\theta}\right) di \\
& - \lambda\left(\int_i \overrightarrow{m_i}^\theta s_{ij} di - \int_i m_i^\star \frac{\partial N_i}{\partial s_j^\theta}\frac{\partial s_{ij}}{\partial N_i} di - \int_i m_i^\star \frac{\partial s_{ij}}{\partial s_j^\theta} di\right),
\end{aligned} \tag{19}
$$

where ELoad_t^{SB} denotes the excess load in time t evaluated at the nationally-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 non-additional households, holding the number and size of installations constant.

Plugging this into (19) and setting the derivative equal to 0 yields

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

This can be rewritten as

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

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 (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 cost 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 cost 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 $\frac{\partial M_j}{\partial s_j^\theta}$, $\overrightarrow{\Delta D}_j^{\theta,\text{ext}}$, $\overrightarrow{s}_j^{\theta,\text{ext}}$, $\frac{\partial N_j}{\partial s_j^\theta}\Big|_{M_j^{\text{st}}}$ $\overrightarrow{\Delta D}_j^{\theta,\text{int}}$, 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 and 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:

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

where $s_{ij} = s_j^{\text{Panel}} N_i^\star + s_j^{\text{kWh}} A_i N_i^\star + s_j^{\text{Cost}} p_j^{\text{Ins}}\,(N_i^\star)$ 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

$$W = -D\,(\text{ELoad}) - \lambda \left(\sum_j \int_{i \in I_j} s_{ij} m_i^\star di - G \right). \tag{22}$$

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

$$\frac{\partial W}{\partial s_j^\theta} = \int_i \sum_{t=0}^{T} \frac{A_{it}}{(1+r)^t} \left| \frac{\partial D_t\left(\text{ELoad}_t^{MD}\right)}{\partial E_{Rt}^{\text{Solar}}} \right| \left(\overrightarrow{m_i}^\theta N_i^\star + m_i^\star \frac{\partial N_i}{\partial s_j^\theta} \right) di$$
$$- \lambda \left(\int_i \overrightarrow{m_i}^\theta s_{ij} di - \int_i m_i^\star \frac{\partial N_i}{\partial s_j^\theta} \frac{\partial s_{ij}}{\partial N_i} di - \int_i m_i^\star \frac{\partial s_{ij}}{\partial s_j^\theta} di \right), \tag{23}$$

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

As in Online 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 using 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 (24), which gives the first-order condition for each subsidy type in each state j:

$$\underbrace{\frac{\partial M_j}{\partial s_j^\theta} \times \left(\overrightarrow{\Delta D}_j^{\theta,\text{ext}} - \lambda \overrightarrow{s}_j^{\theta,\text{ext}}\right)}_{\text{Extensive Margin}} + \underbrace{\left.\frac{\partial N_j}{\partial s_j^\theta}\right|_{M_j^{\text{st}}} \times \left(\overrightarrow{\Delta D}_j^{\theta,\text{int}} - \lambda \overrightarrow{\frac{\partial s}{\partial N}}_j^{\theta,\text{int}}\right)}_{\text{Intensive Margin}} - \underbrace{\lambda M_j \overline{\frac{\partial s_{ij}}{\partial s_j^\theta}}}_{\text{Mechanical Effect}} = 0. \quad (24)$$

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 non-additional households, which are represented the third term in each of the first-order conditions ("Mechanical Effect"). For the damage-minimizing planner, increased subsidies for these non-additional households entail a fiscal cost with no additional decrease in damages. Therefore, the number of non-additional 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 non-additional households. Therefore, each additional dollar of subsidies for a non-additional household is valued at $(1-\lambda)$, reflecting both this increase in utility and the fiscal cost.

C Results Appendix: For Online Publication

C.1 Installation Prices

Table A3 shows the results for estimating solar system installation prices using the Tracking the Sun data using the following regression,

$$p_{ijt}^{\text{Ins}} = p_{R(j)}^{0,\text{Ins}} + p_{R(j)}^{1,\text{Ins}} N_{ijt} + p_{R(j)}^{2,\text{Ins}} (t - 2017) + \varepsilon_{ijt}, \quad (25)$$

where p_{ijt}^{Ins} is the installation cost paid by household i in state j and year t, N_{ijt} is the number panels in the installation, $p_{R(j)}^{0,\text{Ins}}$ is a fixed cost, $p_{R(j)}^{1,\text{Ins}}$ is a per-panel cost, $p_{R(j)}^{2,\text{Ins}}$ is a linear time trend for year t, 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 in 2017, and the coefficient on the number of panels is the per-panel cost in 2017. The linear model is a good fit for the data, as seen in Figure A6, which shows our fitted line against a flexible smoothing function for each region.

Table A3: Solar system installation prices

Dependent Variable:	Total Cost				
Census Region	Full sample	Midwest	Northeast	South	West
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
Constant	5,899.5***	4,784.8***	1,825.2***	8,626.4***	7,048.1***
	(32.9)	(634.1)	(62.0)	(375.3)	(39.6)
Number of Panels	918.1***	1,060.0***	1,062.1***	703.0***	863.8***
	(1.6)	(26.3)	(2.6)	(14.9)	(2.1)
Year (Relative to 2017)	-577.5***	-2,806.6***	-1,165.5***	630.1***	-338.3***
	(9.2)	(192.6)	(15.3)	(77.9)	(11.2)
Fit statistics					
Observations	720,665	1,097	171,306	12,354	535,908
R^2	0.53	0.61	0.71	0.46	0.44
Adjusted R^2	0.53	0.61	0.71	0.46	0.44

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*
Sample limited to between 2014 and 2018, prices measured in 2014 dollars.

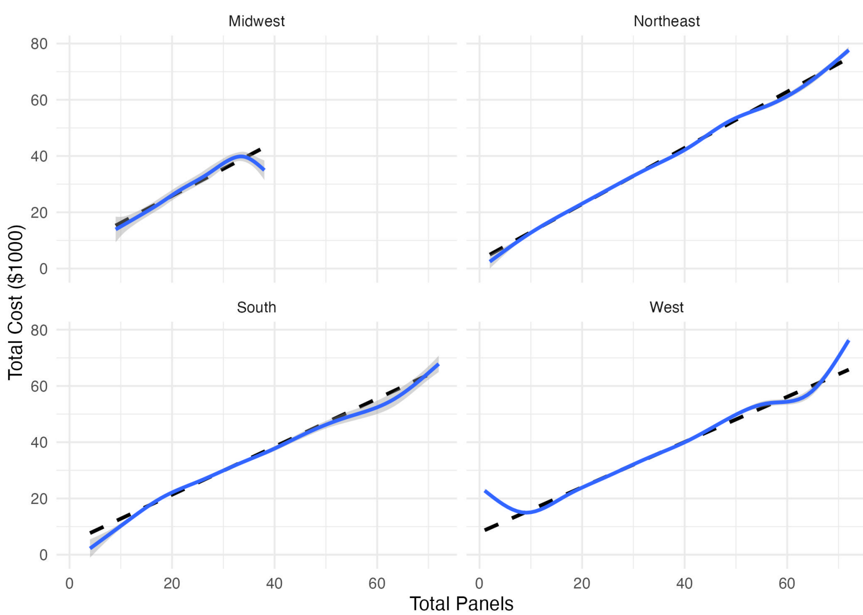


Figure A6: Estimation results for solar system price regression, where the dashed black line is our estimated model in 2017 and the solid blue line shows the fit of a generalized additive model for installations in 2017. Installation prices are measured in 2014 dollars.

	(1)	(2)	(3)	(4)	(5)	(6)
Monetary Benefits	0.106***	0.105***	0.0818***	0.0850***	0.0908***	0.0938***
	(0.0353)	(0.0350)	(0.0166)	(0.0196)	(0.0162)	(0.0182)
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 A4: Regression of log installations on the net present value of total monetary benefits associated with solar panel installations. Monetary benefits measured in thousands of 2014 dollars. Standard errors clustered by state.

C.2 Relationship Between Installations and Monetary Incentives

Table A4 regresses tract-level log installations on the 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. Specifications with "Demographic Controls" include controls for tract-level college completion percentage and percent of voters who voted Democrat in the 2016 presidential election. 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 8.2% to 10.6% increase in installations.

C.3 Installation Size Regressions

Table A5 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 0.07 to 0.11 increase in average panels per installation.

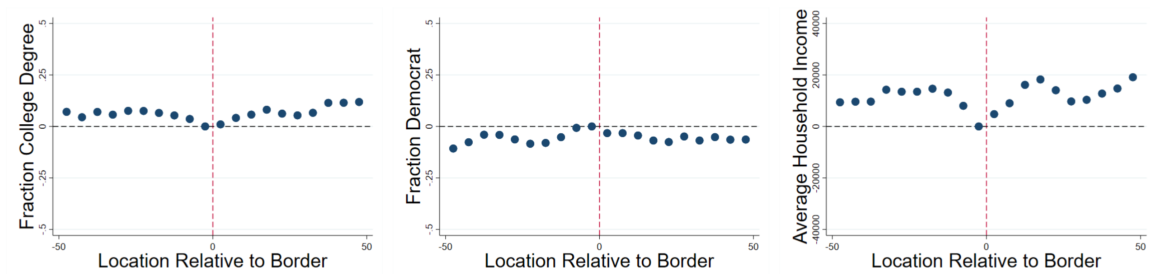
C.4 Border Discontinuities in Household Characteristics

Here, we look for evidence of sorting based on preferences for solar panels on either side of state borders. For each graph, we regress the variable in question on state-border fixed effects and dummy variables for these locations bins and plot these estimated location fixed effects. Figures A7a, A7b, and A7c plot these fixed effects for percent with college degree, percent of voters who voted Democrat in the 2016 presidential election, and average household income,

	(1)	(2)	(3)	(4)	(5)	(6)
Monetary Benefits	0.0737***	0.0856***	0.0787***	0.0878***	0.115***	0.117***
	(0.0222)	(0.0217)	(0.0228)	(0.0221)	(0.0316)	(0.0321)
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 A5: Regression of average panel size on the total monetary benefits associated with solar panel installations. Monetary benefits measured in thousands of 2014 dollars. Standard errors clustered by state.



(a) Percent College Degree (b) Percent Democrat (c) Average Household Income

Figure A7: Border Discontinuities in Household Characteristics. For each graph, we regress the variable in question on state-border fixed effects and dummy variables for these locations bins and plot these estimated location fixed effects.

respectively. There is no clear pattern in these characteristics on either side of the border.

C.5 Border Regressions with Controls for Non-Subsidy Incentives

In order to measure the prevalence of non-subsidy programs aimed at increasing solar adoption, we collect data from DSIRE on the number of state-implemented policies for solar photovoltaics in the residential sector for each state. We categorized non-subsidy programs into three categories: 1) "Financing Programs," policies providing loans and financing options to pay for solar installations, 2) "Access Rules," policies that protect households' ability to install solar panels, and 3) "Building Incentives," programs which create incentives for new building projects to include rooftop solar.[61] We calculated the number of distinct policies in each of these three categories and added these policy counts as controls in our border discontinuity regressions.

[61] "Financing Programs" are programs classified as "Loan Program" and "PACE Financing" in DSIRE, "Access Rules" are programs classified as "Solar/Wind Access Policy" and "Solar/Wind Permitting Standards" which target solar photovoltaics, and "Building Incentives" are programs classified as "Building Energy Codes" and "Green Building Incentives".

	Dependent Variable: Log Installations per Capita			
	(1)	(2)	(3)	(4)
NPV Subsidies ($1000s)	0.0826***	0.105***	0.0854***	0.0810***
	(0.00761)	(0.00882)	(0.0142)	(0.0132)
Financing Programs		0.328***	0.302***	0.307***
		(0.0641)	(0.0605)	(0.0611)
Access Rules			0.289*	0.271
			(0.163)	(0.166)
Building Incentives				0.208*
				(0.110)
				(0.104)
Demographic Controls	YES	YES	YES	YES
Border FE	YES	YES	YES	YES
Distance Bandwidth	10	10	10	10
Tax Controls	Yes	Yes	Yes	Yes
Observations	6,052	6,052	6,052	6,052

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

Table A6: Regression of log installations per capita on the net present value of subsidies for a 15-panel installation within 10 miles of state borders with controls for non-subsidy programs. Subsidies are measured in thousands of 2014 dollars. State-clustered standard errors in parentheses. All regressions contain border fixed effects, tract-level demographic controls, and controls for state tax rates. Non-subsidy policy variables are measured as the number of policies of a given type in the state.

Table A6 shows the results. The first column shows the border discontinuity regression without these additional policy controls, column (2) adds a control variable for the number of Financing Programs in the states, column (3) adds the number of Access Rules in the state, column (4) add the number of Building Incentives. In all regressions, our main coefficient estimating the elasticity of installations with respect to subsidies is very similar to the baseline results, and all of the non-subsidy programs have the expected positive sign.

In Table A7, we repeat this analysis but replace the count of policies in each program type with a dummy equal to one if the state has any programs of that type. The results are very similar to those in A6.

C.6 Power Plant Model

Here we present additional information about the power plant model estimation results. Figures A8 and A9 show the model fit graphs for the main paper, broken out into regions and evaluated on out-of-sample data from 2020. These show that model performance is consistent within each interconnection.

	Dependent Variable: Log Installations per Capita			
	(1)	(2)	(3)	(4)
NPV Subsidies ($1000s)	0.0826***	0.107***	0.103***	0.0933***
	(0.00761)	(0.00984)	(0.0131)	(0.0137)
Has Financing Programs		0.587***	0.563***	0.575***
		(0.148)	(0.125)	(0.129)
Has Access Rules			0.0968	0.0827
			(0.212)	(0.216)
Has Building Incentives				0.308*
				(0.171)
Demographic Controls	YES	YES	YES	YES
Border FE	YES	YES	YES	YES
Distance Bandwidth	10	10	10	10
Tax Controls	Yes	Yes	Yes	Yes
Observations	6,052	6,052	6,052	6,052

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

Table A7: Regression of log installations per capita on the net present value of subsidies for a 15-panel installation within 10 miles of state borders with controls for non-subsidy programs. Subsidies are measured in thousands of 2014 dollars. State-clustered standard errors in parentheses. All regressions contain border fixed effects, tract-level demographic controls, and controls for state tax rates. Non-subsidy policy variables are dummy variables equal to one if the state has at least one policy of a given type.

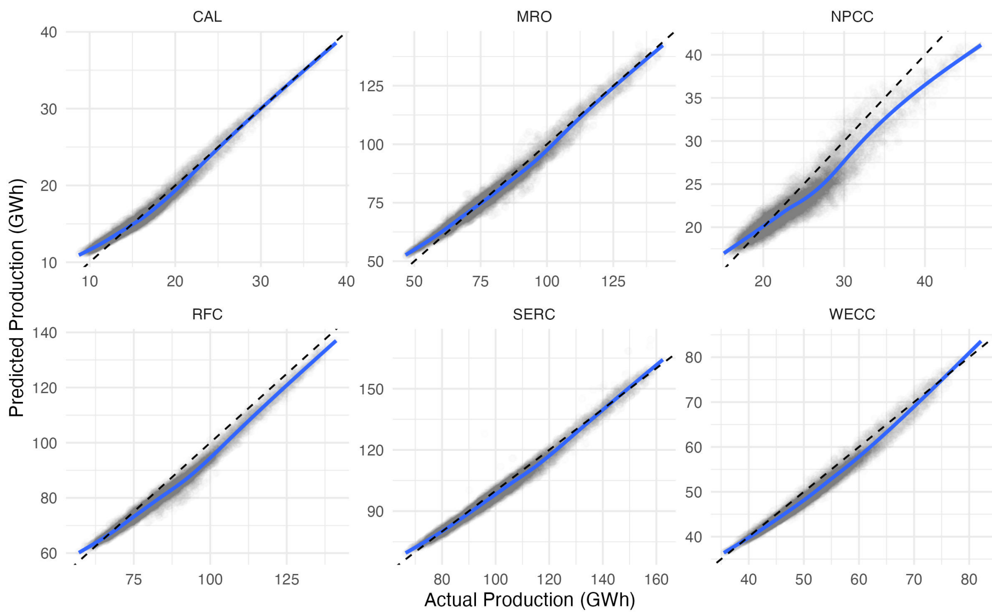


Figure A8: Model fit at the region level in 2020, excluding Texas as there is only one region in the Texas interconnection. Dots represent an hour of production for each region, smoothed lines show the fit of a generalized additive model.

	(1)	(2)	(3)
		State-Specific Subsidies	
	Baseline	Welfare Max	Damage Min
Unit Subsidies	7.1	0.8	1.0
Cost Subsidies	80.9	1.5	0.8
kWh Subsidies	8.4	97.7	98.2
Total	100.0	100.0	100.0

Table A8: 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.

A10 shows how fuel mix varies for each region, using in-sample 2019 data.

C.7 Type of Subsidy in Nationally-Optimal System

We now analyze how the government should optimally allocate subsidies across the three subsidy types: cost-based subsidies, per-panel 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.[62] We then average this hypothetical subsidy value over all households. Table A8 shows the percent of the total subsidy value coming from each subsidy type in each simulation. Under the current system,

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

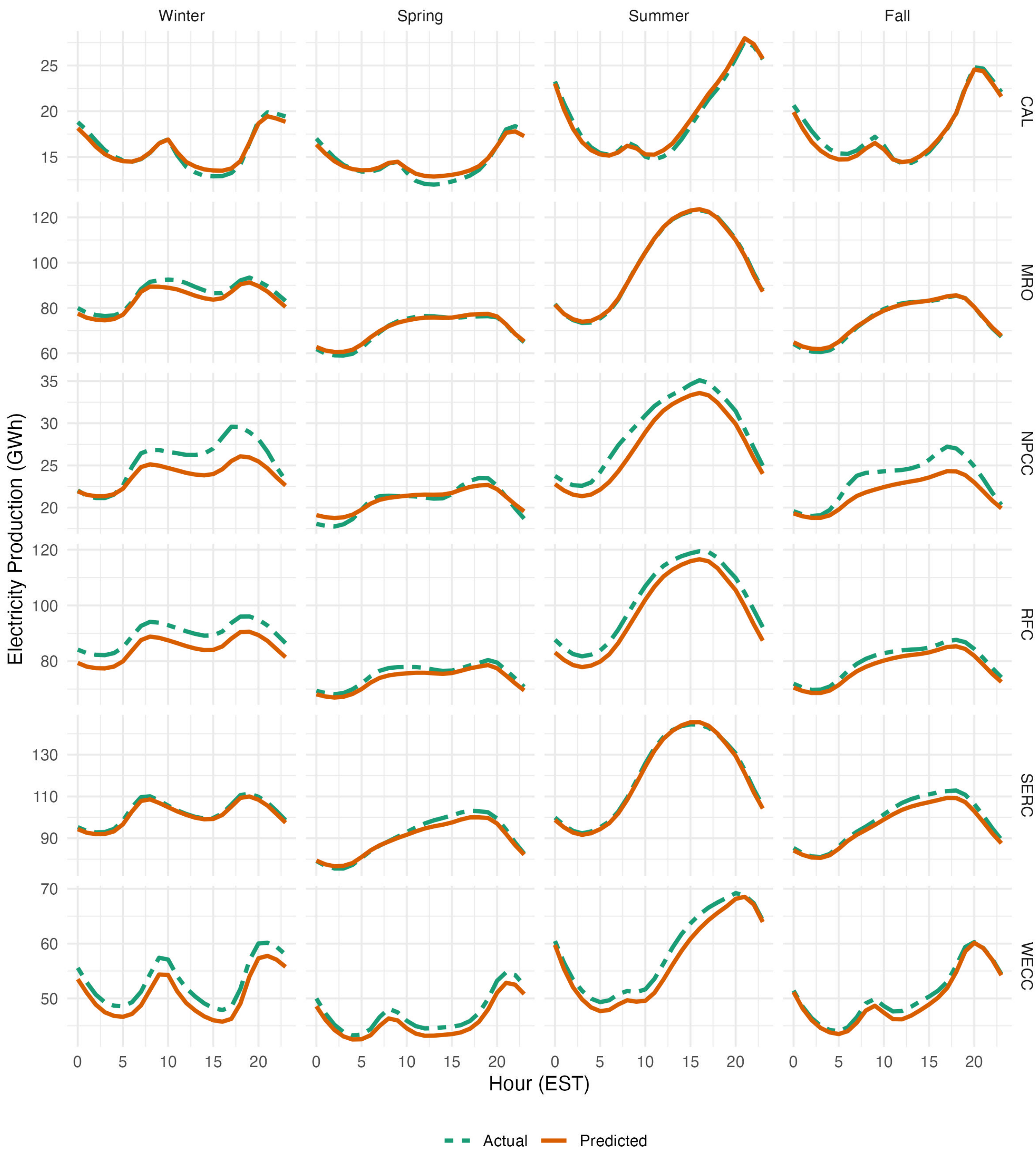


Figure A9: Model fit at the region level by hour and season in 2020. The dashed green line gives electricity production in the data while the solid orange line gives predicted production.

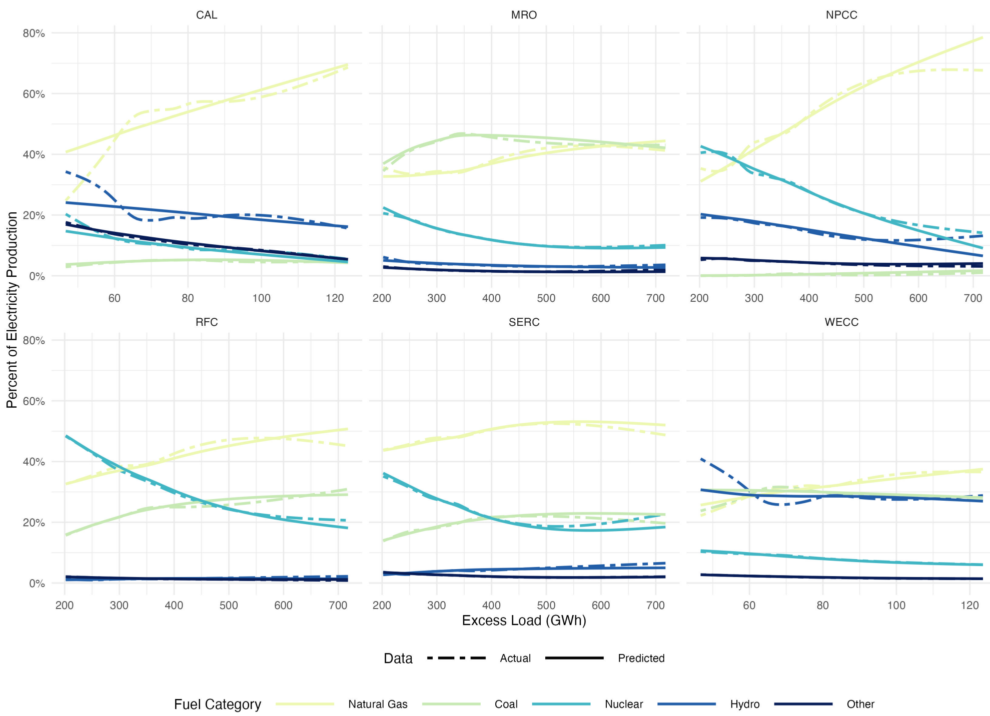


Figure A10: Fuel mix of production by region. The X-axis gives excess load at the interconnection level and the Y-axis gives the percent of electricity production that is produced by each of the fuel types using data from 2020. The dashed lines show the fuel mix in the data while the solid lines show the simulated fuel mix.

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. Intuitively, production-based subsidies incentivize installations for households where sunlight, and therefore environmental benefits, are high.[63] However, as we show in Section 6.5, the gains to reallocating across subsidy types within states are small relative to the gains from reallocating subsidies *across* states.

C.8 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.

[63]One important caveat is that we assume households' discount rate is given by the inverse of the real interest rate.

	(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	14.7	14.9	15.0	14.9	15.0
Northeast	15.1	14.9	15.0	14.9	15.0
South	14.8	15.1	15.1	15.1	15.1
West	15.0	15.2	15.1	15.2	15.1

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

C.9 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 simulated number of solar panel installations per 1000 households given the current subsidies and under the damage-minimizing subsidies.

C.10 Unconstrained Nationally-Optimal Subsidies with $\lambda = 1$

Theory We define welfare as

$$W = \underbrace{\int_i V_i di}_{\text{Utility}} - \underbrace{D\left(\text{ELoad}\right)}_{\text{Damages}} - \underbrace{\sum_j \int_{i \in I_j} s_{ij} m_i^\star di}_{\text{Government Cost}}. \tag{26}$$

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

$$\underbrace{\frac{\partial M_j}{\partial s_j^\theta} \times \left(\overrightarrow{\Delta D}_j^{\theta,\text{ext}} - \overrightarrow{s}_j^{\theta,\text{ext}}\right)}_{\text{Extensive Margin}} + \underbrace{\left.\frac{\partial N_j}{\partial s_j^\theta}\right|_{M_j^{\text{st}}} \times \left(\overrightarrow{\Delta D}_j^{\theta,\text{int}} - \overrightarrow{\frac{\partial s}{\partial N}}_j^{\theta,\text{int}}\right)}_{\text{Intensive Margin}} = 0, \tag{27}$$

where all objects are as defined in Section 2.3.

The optimal policy balances two forces: 1) the decrease in damages and 2) the increase in cost due to an increase in the number of panel installed, through both extensive and intensive margin adjustments. Importantly, note that household utility does not show up in this formula. This is because there is no first-order welfare effect on households for marginal

	Expected Subsidy		Installs/1000HH			Expected Subsidy		Installs/1000HH	
	Baseline	Optimal	Baseline	Optimal		Baseline	Optimal	Baseline	Optimal
Alabama	5.8	15.1	3.3	7.2	Nebraska	7.6	13.8	4.1	6.9
Arizona	11.2	14.0	8.7	11.1	Nevada	6.6	13.9	4.8	9.0
Arkansas	5.8	13.9	3.3	6.6	New Hampshire	18.1	9.4	16.7	8.0
California	11.7	13.0	15.7	17.7	New Jersey	27.6	17.7	38.7	16.8
Colorado	9.4	12.6	6.2	8.1	New Mexico	10.6	14.4	8.3	11.5
Connecticut	15.7	9.5	18.3	10.9	New York	17.1	9.5	16.2	8.6
Delaware	10.0	18.2	7.0	14.2	North Carolina	8.9	15.1	5.7	9.7
Florida	10.3	16.0	7.2	11.8	North Dakota	10.0	12.6	3.8	4.8
Georgia	5.8	15.1	4.8	10.6	Ohio	6.6	17.2	3.6	8.7
Idaho	7.4	11.7	3.3	4.7	Oklahoma	5.8	14.7	3.3	7.1
Illinois	12.5	17.1	7.0	10.2	Oregon	11.1	9.9	6.2	5.5
Indiana	9.6	17.7	4.4	8.7	Pennsylvania	5.7	17.5	4.2	11.1
Iowa	16.9	13.2	8.0	5.8	Rhode Island	15.9	9.6	16.4	9.6
Kansas	11.7	14.2	6.6	8.1	South Carolina	9.3	15.3	4.7	8.7
Kentucky	6.3	14.3	3.6	7.2	South Dakota	11.3	12.9	4.0	4.5
Louisiana	15.2	14.2	8.2	7.6	Tennessee	7.6	14.7	3.9	7.1
Maine	5.3	9.5	4.7	6.7	Texas	16.0	12.1	8.9	6.4
Maryland	12.5	18.3	10.3	16.8	Utah	8.0	12.3	3.6	5.4
Massachusetts	24.7	9.5	36.6	10.4	Vermont	12.8	9.2	14.8	11.0
Michigan	6.2	12.5	3.7	6.3	Virginia	5.8	15.3	4.1	9.2
Minnesota	12.1	12.8	6.3	6.6	Washington	18.8	9.4	10.3	4.6
Mississippi	5.8	14.9	3.5	7.6	West Virginia	5.8	17.6	2.6	7.1
Missouri	9.4	13.8	5.2	7.4	Wisconsin	16.9	16.0	9.2	8.5
Montana	9.0	10.8	4.1	4.7	Wyoming	6.0	12.1	2.4	4.1

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 2014 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.

	Expected Subsidy		Installs/1000HH			Expected Subsidy		Installs/1000HH	
	Baseline	Optimal	Baseline	Optimal		Baseline	Optimal	Baseline	Optimal
Alabama	5.8	15.6	3.3	7.5	Nebraska	7.6	12.9	4.1	6.3
Arizona	11.2	13.4	8.7	10.5	Nevada	6.6	13.1	4.8	8.4
Arkansas	5.8	13.2	3.3	6.2	New Hampshire	18.1	3.9	16.7	5.0
California	11.7	11.3	15.7	15.3	New Jersey	27.6	20.6	38.7	21.4
Colorado	9.4	10.5	6.2	6.8	New Mexico	10.6	14.1	8.3	11.2
Connecticut	15.7	4.1	18.3	6.9	New York	17.1	4.1	16.2	5.4
Delaware	10.0	21.8	7.0	19.2	North Carolina	8.9	15.5	5.7	10.0
Florida	10.3	17.3	7.2	13.3	North Dakota	10.0	10.5	3.8	4.0
Georgia	5.8	15.5	4.8	11.0	Ohio	6.6	19.7	3.6	10.8
Idaho	7.4	8.7	3.3	3.6	Oklahoma	5.8	14.7	3.3	7.1
Illinois	12.5	19.3	7.0	12.4	Oregon	11.1	5.0	6.2	3.7
Indiana	9.6	20.8	4.4	11.4	Pennsylvania	5.7	20.2	4.2	14.0
Iowa	16.9	11.5	8.0	5.1	Rhode Island	15.9	4.1	16.4	6.1
Kansas	11.7	13.7	6.6	7.7	South Carolina	9.3	16.0	4.7	9.3
Kentucky	6.3	13.9	3.6	6.9	South Dakota	11.3	10.8	4.0	3.8
Louisiana	15.2	13.8	8.2	7.3	Tennessee	7.6	14.6	3.9	7.1
Maine	5.3	4.1	4.7	4.2	Texas	16.0	9.6	8.9	5.1
Maryland	12.5	21.9	10.3	22.7	Utah	8.0	9.9	3.6	4.4
Massachusetts	24.7	3.9	36.6	6.6	Vermont	12.8	3.4	14.8	6.7
Michigan	6.2	10.0	3.7	5.1	Virginia	5.8	15.9	4.1	9.7
Minnesota	12.1	10.6	6.3	5.5	Washington	18.8	4.2	10.3	3.0
Mississippi	5.8	15.1	3.5	7.7	West Virginia	5.8	20.6	2.6	9.2
Missouri	9.4	12.7	5.2	6.8	Wisconsin	16.9	17.1	9.2	9.4
Montana	9.0	6.7	4.1	3.3	Wyoming	6.0	9.5	2.4	3.2

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 2014 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.

households (i.e. households who choose to install solar panels in response to the increase in subsidies) because of the envelope theorem. Further, the utility increase for non-additional households (i.e. households who already chose to install solar panels before the increase in subsidies) associated with receiving a larger subsidy for existing panels is exactly offset by the cost of increasing subsidies for these households.[64]

Results Table A12 presents the baseline and nationally-optimal subsidy in each state. Nationally-optimal subsidies are lowest in Vermont, at $6,600. Nationally-optimal subsidies are over twice as high in most of the Mid-Atlantic, with the highest subsidies at over $14,400 in Delaware and Maryland.

D Extensions and Robustness Appendix: For Online Publication

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 and in (3) we also add the percent of voters who voted Democrat in the 2016 presidential election. To these variables, (4) adds home-ownership rate rate and (5) instead adds average income. Note that Column (3) is the same as our baseline specification. The results are qualitatively very similar across specifications.

D.2 Alternative Discounting of Future Subsidy Payments

The optimal subsidy results where households discount future subsidy payments at an implicit interest rate of 15% are summarized in Table A14. In all three specifications, the distribution of subsidies and installations across space is similar to those in the baseline model.

[64] This is a direct consequence of 1) quasilinear utility and a utilitarian welfare function with equal Pareto weights, which together imply that marginal social welfare weights (Saez and Stantcheva, 2016) are equalized and there are no effects of total welfare of wealth redistribution, and 2) the assumption of that the marginal cost of public funds is equal to one: the social planner values an increase in consumption for a given household the same as an increase in government revenue.

	Expected Subsidy			Expected Subsidy	
	Baseline	Optimal		Baseline	Optimal
Alabama	5.8	11.7	Nebraska	7.6	10.6
Arizona	11.2	10.7	Nevada	6.6	10.6
Arkansas	5.8	10.7	New Hampshire	18.1	6.7
California	11.7	9.8	New Jersey	27.6	13.9
Colorado	9.4	9.5	New Mexico	10.6	11.1
Connecticut	15.7	6.8	New York	17.1	6.8
Delaware	10.0	14.4	North Carolina	8.9	11.7
Florida	10.3	12.4	North Dakota	10.0	9.6
Georgia	5.8	11.7	Ohio	6.6	13.5
Idaho	7.4	8.7	Oklahoma	5.8	11.3
Illinois	12.5	13.3	Oregon	11.1	7.1
Indiana	9.6	14.0	Pennsylvania	5.7	13.8
Iowa	16.9	10.0	Rhode Island	15.9	6.8
Kansas	11.7	10.9	South Carolina	9.3	11.9
Kentucky	6.3	11.0	South Dakota	11.3	9.7
Louisiana	15.2	10.9	Tennessee	7.6	11.3
Maine	5.3	6.8	Texas	16.0	9.1
Maryland	12.5	14.5	Utah	8.0	9.3
Massachusetts	24.7	6.8	Vermont	12.8	6.6
Michigan	6.2	9.4	Virginia	5.8	11.9
Minnesota	12.1	9.6	Washington	18.8	6.8
Mississippi	5.8	11.5	West Virginia	5.8	13.9
Missouri	9.4	10.5	Wisconsin	16.9	12.5
Montana	9.0	7.9	Wyoming	6.0	9.1

Table A12: The first two columns shows the state-level subsidies given the current system and the unconstrained nationally-optimal subsidies. Subsidies are measured as the average present discounted value of subsidies for a 15-panel installation, measured in thousands of 2014 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.

	(1)	(2)	(3)	(4)	(5)
I. Δ Average Subsidy ($Thousands)					
Midwest	2.6	2.6	2.7	2.7	2.7
Northeast	-8.3	-8.6	-9.1	-9.1	-9.3
South	2.4	2.4	2.4	2.5	2.4
West	1.0	1.1	1.2	1.2	1.4
II. Δ Installations per 1000HHs					
Midwest	5.1	5.2	5.4	5.4	5.5
Northeast	-4.3	-4.2	-4.0	-4.1	-3.9
South	3.7	3.8	3.9	3.9	4.1
West	0.7	0.8	0.9	0.9	1.0
III. Δ Annual Damages Offset ($Millions)					
Total	29.4	30.0	30.3	30.3	30.1
Nonpecuniary Component Depends On:					
College Share	No	Yes	Yes	Yes	Yes
Percent Democrat	No	No	Yes	Yes	Yes
Homeowner Share	No	No	No	Yes	No
Average Income	No	No	No	No	Yes

Table A13: Counterfactual results under alternative model specifications. Each entry shows the change of moving from the current system of subsidies to the nationally-optimal cost-neutral 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. All monetary values are measured in 2014 dollars. See text for details on each model specification. "Percent Democrat" refers to the percent of voters who voted Democrat in the 2016 presidential election.

	(1)	(2)	(3)	(4)
		State-Specific Subsidies		
	Baseline	Welfare Max	Damage Min	Unconstrained
I. Average Subsidy ($Thousands)				
Midwest	10.0	15.4	15.7	11.7
Northeast	17.1	13.1	11.7	9.9
South	10.6	14.6	14.5	11.2
West	11.7	12.6	10.7	9.6
II. Installations per 1000HHs				
Midwest	5.3	8.2	8.8	5.9
Northeast	20.7	11.4	11.3	8.6
South	6.5	9.1	9.3	6.7
West	11.9	13.2	11.5	10.1
National	9.9	10.4	10.1	7.8
III. Annual Damages Offset ($Millions)				
Total	263.6	293.6	296.5	217.0
IV. Annuitized Total Subsidies Paid ($Millions)				
	383.3	383.3	383.3	217.0

Table A14: Nationally-optimal subsidies with alternative household discounting. 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 third panel gives the total damages offset by rooftop solar. The final panel gives total government cost under each subsidy scheme converted to an annuity value. “Unconstrained” refers to the optimal unconstrained subsidies with the marginal value of public funds equal to one ($\lambda = 1$). All monetary values are measured in 2014 dollars.

	(1)	(2)	(3)	
	Current	State-Specific Optimal Subsidies		
	Subsidies	Welfare Max	Damage Min	Unconstrained
Unit Subsidies	7.1	16.9	16.4	10.4
Cost Subsidies	80.9	83.1	83.6	89.6
kWh Subsidies	8.4	0.0	0.0	0.0
Total	100.0	100.0	100.0	100.0

Table A15: Percent of total subsidy value from each type of subsidy for a 15-panel installation averaged across all households in the model with alternative household discounting. Each column shows the subsidy values for a different simulation. “Unconstrained” refers to the optimal unconstrained subsidies with the marginal value of public funds equal to one ($\lambda = 1$).

Table A15 shows how the government should optimally allocate subsidies across the three subsidy types: cost-based subsidies, per-panel subsidies, and production-based (per-kWh) subsidies. For this, we calculate the subsidy value every household in the model would receive if they purchased a 15-panel installation and then average this hypothetical subsidy value over all households. We then calculate the percent of the total subsidy value coming from each subsidy type in each simulation. In all of the optimal subsidy systems, the government relies entirely on upfront subsidies. This makes sense, as production subsidies are not as highly valued, per dollar, as these other two subsidies types.

Table A16 shows the damages offset per additional dollar of government cost associated with marginal subsidy increases around the current system of subsidies. As before, we calculate this by first simulating the model 1) under the current system of subsidies and 2) with marginally more generous subsidies of a given type in a given state. We then calculate the damages offset per dollar of this particular subsidy as the difference in damages between the two simulations divided by the difference in the fiscal cost. We repeat this process for each subsidy type in each state.

We can see that production subsidies are much less cost-effective than investment and per-panel subsidies. This makes sense, as households in this model heavily discount subsidies received in the future.

D.3 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 plus a factor proportional to the square of flow on the line: $L_{Rt} = \alpha_{1R} + \alpha_{2R}\left(\text{Load}_{Rt} - E_{Rt}^{\text{Solar}}\right)^2$. 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, $\text{ELoad}_{Rt}^{LL} = \text{Load}_{Rt} - E_{Rt}^{\text{NonD}} - E_{Rt}^{\text{Solar}} + L_{Rt}$. 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

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

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

	Subsidy Type				Subsidy Type		
	Panel	Cost	kWh		Panel	Cost	kWh
Alabama	0.68	0.68	0.32	Nebraska	0.56	0.55	0.28
Arizona	0.41	0.41	0.24	Nevada	0.60	0.59	0.29
Arkansas	0.62	0.62	0.30	New Hampshire	0.23	0.23	0.14
California	0.43	0.43	0.24	New Jersey	0.36	0.36	0.24
Colorado	0.46	0.46	0.24	New Mexico	0.51	0.50	0.27
Connecticut	0.25	0.25	0.15	New York	0.24	0.24	0.14
Delaware	0.68	0.67	0.36	North Carolina	0.58	0.57	0.30
Florida	0.58	0.58	0.31	North Dakota	0.45	0.45	0.24
Georgia	0.69	0.68	0.32	Ohio	0.75	0.75	0.36
Idaho	0.47	0.47	0.23	Oklahoma	0.67	0.66	0.32
Illinois	0.56	0.56	0.31	Oregon	0.27	0.27	0.16
Indiana	0.67	0.66	0.35	Pennsylvania	0.80	0.79	0.38
Iowa	0.36	0.35	0.21	Rhode Island	0.25	0.25	0.15
Kansas	0.47	0.47	0.26	South Carolina	0.55	0.54	0.29
Kentucky	0.39	0.39	0.24	South Dakota	0.43	0.43	0.23
Louisiana	0.42	0.41	0.24	Tennessee	0.60	0.59	0.30
Maine	0.40	0.40	0.19	Texas	0.34	0.33	0.20
Maryland	0.61	0.60	0.34	Utah	0.42	0.41	0.23
Massachusetts	0.18	0.18	0.12	Vermont	0.27	0.27	0.15
Michigan	0.53	0.53	0.25	Virginia	0.69	0.68	0.33
Minnesota	0.41	0.41	0.23	Washington	0.22	0.22	0.14
Mississippi	0.67	0.66	0.32	West Virginia	0.81	0.80	0.38
Missouri	0.50	0.50	0.26	Wisconsin	0.44	0.44	0.26
Montana	0.39	0.38	0.20	Wyoming	0.52	0.52	0.25

Table A16: Damages offset per additional dollar of government funds associated with marginal subsidy increases around the current system of subsidies with alternative household discounting. The first column ("Panel") gives the damages offset per dollar associated with marginal increases in panel-based subsidies, s_j^{Panel}, 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}.

an additional benefit from offsetting line losses, captured by the term $2\alpha_{2R}\left(\text{Load}_{Rt} - E_{Rt}^{\text{Solar}}\right)$, which is the marginal change in losses. Including line losses increase the average damages offset by marginal installers, $\overrightarrow{\Delta D}_j^{\theta,\text{ext}}$, and by intensive margin installers, $\overrightarrow{\Delta D}_j^{\theta,\text{int}}$, when calculating nationally-optimal subsidies.[65]

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 $\alpha_1 = 0.25\gamma_R \sum_t \left(\text{Load}_{Rt} - E_{Rt}^{\text{Solar}}\right)$ and $\alpha_2 = 0.75\gamma_R \frac{\sum_t\left(\text{Load}_{Rt} - E_{Rt}^{\text{Solar}}\right)}{\sum_t\left(\text{Load}_{Rt} - E_{Rt}^{\text{Solar}}\right)^2}$.

Results The results are summarized in Table A17. The first column gives cost-neutral 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, (4) state-specific unconstrained nationally-optimal subsidies. In all three 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 larger than in the baseline model.

D.4 Transmission Constraints

The results in the model with transmission constraints are summarized in Table A17. The first column gives cost-neutral subsidies, installations, and damages offset given the current system of subsidies when we account for transmission constraints. The annual damages offset are slightly larger than the baseline model in which we do not account for line losses.

The following columns summarize the results under (2) state-specific welfare-maximizing subsidies, (3) state-specific damage-minimizing subsidies, (4) state-specific unconstrained nationally-optimal subsidies. We find that current spending on subsidies exceeds the optimal amount by roughly 45% in this case.

[65] 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.

	(1)	(2)	(3)	(4)
		State-Specific Subsidies		
	Baseline	Welfare Max	Damage Min	Unconstrained
I. Average Subsidy ($Thousands)				
Midwest	10.0	15.4	16.1	13.1
Northeast	17.1	13.1	11.6	10.9
South	10.6	14.5	14.2	12.2
West	11.7	12.5	10.8	10.4
II. Installations per 1000HHs				
Midwest	5.4	8.2	9.0	6.7
Northeast	20.3	11.3	11.2	9.4
South	6.6	8.9	9.1	7.4
West	11.9	13.1	11.6	11.0
National	9.9	10.3	10.1	8.5
III. Annual Damages Offset ($Millions)				
Total	288.5	321.6	324.0	264.2
IV. Annuitized Total Subsidies Paid ($Millions)				
	380.8	380.8	380.8	263.9

Table A17: Nationally-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 third panel gives the total damages offset by rooftop solar. The final panel gives total government cost under each subsidy scheme converted to an annuity value. "Unconstrained" refers to the optimal unconstrained subsidies with the marginal value of public funds equal to one ($\lambda = 1$). All monetary values are measured in 2014 dollars.

	(1)	(2)	(3)	(4)
		State-Specific Subsidies		
	Baseline	Welfare Max	Damage Min	Unconstrained
I. Average Subsidy ($Thousands)				
Midwest	10.0	14.4	14.4	12.0
Northeast	17.1	15.5	16.6	13.1
South	10.6	12.8	11.4	10.6
West	11.7	13.6	13.1	11.4
II. Installations per 1000HHs				
Midwest	5.4	7.4	7.5	6.1
Northeast	20.3	13.7	15.1	11.1
South	6.6	7.6	6.9	6.3
West	11.9	14.4	14.2	11.8
National	9.9	10.4	10.3	8.5
III. Annual Damages Offset ($Millions)				
Total	299.1	311.7	313.5	256.3
IV. Annuitized Total Subsidies Paid ($Millions)				
	380.8	380.8	380.8	259.7

Table A18: Nationally-optimal subsidies when accounting for transmission constraints. 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 third panel gives the total damages offset by rooftop solar. The final panel gives total government cost under each subsidy scheme converted to an annuity value. "Unconstrained" refers to the optimal unconstrained subsidies with the marginal value of public funds equal to one ($\lambda = 1$). All monetary values are measured in 2014 dollars.

D.5 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 nationally-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.[66] 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.[67]

Results We calculate the environmental benefits of this improved storage technology and the nationally-optimal subsidies given the new storage technology for three values of ω in Table A19. 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 $125 million annually. Column (3) recalculates the nationally-optimal cost-neutral subsidies given that the new storage technology is in place. We find that the nationally-optimal subsidies are very similar to the baseline case and that implementing the

[66]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.

[67]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.

nationally-optimal subsidies leads to similar reductions in environmental damages as we find without the improved storage technology. Column (4) shows the unconstrained nationally-optimal subsidies given the new storage technology. The nationally-optimal unconstrained subsidies are very similar to the baseline case.

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 nationally-optimal subsidies and the environmental benefits associated with implementing those subsidies are similar to those in the baseline case.

D.6 Cleaner Electricity Production

We present our results when we allow changes in electricity production in Table A20. 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 cost-neutral 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 welfare-maximizing cost-neutral subsidies given the baseline technology leads to an increase in environmental benefits of 6.2%.

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 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.[68] We refer to their reference case projection as the mid-cost projection. Across the three scenarios, we find that moving to the welfare-maximizing cost-neutral subsidies leads to a 17-20% increase in aggregate environmental benefits.

In the fifth 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 welfare-maximizing cost-neutral subsidies leads to a 11.6% increase in aggregate environmental benefits in this case.

Columns 6 through 10 repeat this exercise with unconstrained nationally-optimal subsidies. In all cases, current subsidies are overfunded relative to the optimum. Moving to unconstrained nationally-optimal subsidies involves cutting funding for subsidies by 55% to

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
			$\omega = .25$			$\omega = .5$			$\omega = .75$	
	Baseline	Improved Storage	+Optimal Subsidies	+Unconstrained Optimal	Improved Storage	+Optimal Subsidies	+Unconstrained Optimal	Improved Storage	+Optimal Subsidies	+Unconstrained Optimal
I. Average Subsidy ($Thousands)										
Midwest	10.0	10.0	15.2	11.9	10.0	15.1	11.9	10.0	15.1	11.8
Northeast	17.1	17.3	12.9	9.9	17.3	12.9	9.9	17.3	12.8	9.8
South	10.6	11.0	14.4	11.2	11.0	14.4	11.2	11.0	14.3	11.2
West	11.7	12.4	12.8	9.8	12.4	12.9	9.9	12.4	13.0	10.0
II. Installations per 1000HHs										
Midwest	5.4	5.4	8.0	6.0	5.4	7.9	6.0	5.4	7.9	6.0
Northeast	20.3	20.3	11.1	8.5	20.3	11.0	8.5	20.3	11.0	8.5
South	6.6	6.6	8.9	6.7	6.6	8.8	6.7	6.6	8.8	6.7
West	11.9	11.9	13.4	10.4	11.9	13.6	10.5	11.9	13.7	10.6
III. Annual Damages Offset ($Millions)										
Total	262.3	382.1	412.6	340.0	472.8	503.4	431.6	533.3	563.9	492.1
IV. Annuitized Total Subsidies Paid ($Millions)										
National	380.8	380.8	380.8	222.4	380.8	380.8	224.5	380.8	380.8	224.6

Table A19: Nationally-optimal cost-neutral and unconstrained 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 third panel gives the total damages offset by rooftop solar and by the increased storage technology of renewable energy. The final panel gives total government cost under each subsidy scheme converted to an annuity value. "Unconstrained Optimal" refers to the optimal unconstrained subsidies with the marginal value of public funds equal to one ($\lambda = 1$). All monetary values are measured in 2014 dollars.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Cost Neutral					Unconstrained				
	Current	Increased Renewables			Clean	Current	Increased Renewables			Clean
	Tech	High Cost	Mid Cost	Low Cost	Coal	Tech	High Cost	Mid Cost	Low Cost	Coal
I. Δ Average Subsidy ($Thousands)										
Midwest	5.4	6.7	6.9	7.1	3.5	1.9	1.6	1.5	1.3	-2.2
Northeast	-4.0	-3.0	-2.8	-2.7	-4.3	-7.2	-7.6	-7.8	-8.0	-9.9
South	3.9	4.9	5.0	5.0	3.9	0.6	0.2	-0.1	-0.5	-2.1
West	0.9	-1.1	-1.5	-1.7	1.6	-2.2	-5.0	-5.6	-6.0	-4.0
II. Δ Installations per 1000HHs										
Midwest	2.7	3.7	3.8	4.0	1.6	0.7	0.5	0.4	0.3	-1.1
Northeast	-9.1	-8.0	-7.9	-7.7	-9.5	-11.8	-12.0	-12.2	-12.3	-13.5
South	2.4	3.2	3.3	3.3	2.4	0.2	-0.1	-0.3	-0.4	-1.2
West	1.2	-1.4	-1.8	-1.9	2.5	-1.9	-4.4	-4.8	-5.0	-3.1
III. %Δ Environmental Benefits										
Total	11.6	17.4	19.9	20.5	11.6	-16.6	-22.8	-24.0	-25.9	-33.1
IV. %Δ Total Subsidies Paid										
Total	0	0	0	0	0	-42.6	-55.0	-58.0	-60.5	-64.8

Table A20: Nationally-optimal cost-neutral and unconstrained subsidies under alternative assumptions about central generation energy production. Each entry of Columns (1) through (4) shows the change of moving from the current system of subsidies to the welfare-maximizing cost-neutral system of subsidies given the specification in question. Each entry of Columns (5) through (8) shows the change of moving from the current system of subsidies to the unconstrained nationally-optimal system of subsidies given the specification in question. "Unconstrained" refers to the optimal unconstrained subsidies with the marginal value of public funds equal to one ($\lambda = 1$). All monetary values are measured in 2014 dollars. See text for details on each model specification.

65% across specifications.